Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra

Michael H. Neumann[1] and Rainer von Sachs[2]

Abstract. We derive minimax rates for estimation in anisotropic smoothness classes. This rate is attained by a coordinatewise thresholded wavelet estimator based on a tensor product basis with separate scale parameter for every dimension. It is shown that this basis is superior to its one-scale multiresolution analog, if different degrees of smoothness in different directions are present.
As an important application we introduce a new adaptive wavelet estimator of the time-dependent spectrum of a locally stationary time series. Using this model which was recently developed by Dahlhaus, we show that the resulting estimator attains nearly the rate, which is optimal in Gaussian white noise, simultaneously over a wide range of smoothness classes. Moreover, by our new approach we overcome the difficulty of how to choose the right amount of smoothing, i.e. how to adapt to the appropriate resolution, for reconstructing the local structure of the evolutionary spectrum in the time-frequency plane.

1991 Mathematics Subject Classification. Primary 62G07, 62M15; secondary 62E20, 62M10
Key words and phrases. Anisotropic smoothness classes, adaptive estimation, optimal rate of convergence, wavelet thresholding, tensor product basis, time-frequency plane, locally stationary time series, evolutionary spectrum
Short title. Adaptive estimation of evolutionary spectra

[1]Weierstrass Institute for Applied Analysis and Stochastics, Berlin
[2]Department of Mathematics, University of Kaiserslautern

1. Introduction

There is a wide range of fields in which an observed time series shows a nonstationary behavior (by transients, amplitude or frequency modulation, quasi-oscillating behavior, etc.). These can be found, e.g. in many physical phenomena (occurring in geophysics, in transmission problems like radio propagation or in speech and sound analysis), and from economical data analysis, also. A recent approach for modelling certain kinds of these instationarities is by the introduction of the class of locally stationary processes (Dahlhaus, 1993) which both controls the departure from stationarity and gives a frame for asymptotic theory. As in the Cramér representation for stationary processes the spectrum, which now becomes time dependent, controls the evolution of the variance-covariance distribution of the process over frequency and over time.

In the present paper we develop nonlinear wavelet estimators for this kind of time-varying spectral density: With this we address the problem of finding the right amount of smoothing of an estimator which should adaptively reconstruct the underlying structure of the spectrum in the time-frequency plane. Motivated by this problem, we study first a question of more general importance. Inference about the spectrum of a nonstationary time series is a two-dimensional estimation problem with two particular directions, time and frequency, on the plane. If, in this situation and, more generally for any multidimensional curve estimation, the underlying curve shows different degrees of smoothness in the different directions, then the construction of the estimator should properly take this into account. Hence, to establish a benchmark for our estimator we derive first minimax rates for estimation in anisotropic smoothness classes. Because this question is of general interest, we do not assume any specific observation model, but we investigate this problem in Gaussian white noise. For simplicity we consider the two-dimensional case and restrict ourselves to anisotropic Sobolev classes. Straightforward generalizations can be thought of for higher dimensions and other smoothness classes like Hölder and Besov, also. We show that appropriately tuned wavelet estimators are able to attain the optimal rate of convergence in these classes. These estimators use coordinatewise nonlinear thresholding of empirical wavelet coefficients. The rate for the risk can be easily found by analyzing a certain complexity functional, which describes the amount of data compression of a basis in a given smoothness class. We show that we obtain a suitable higher-dimensional basis by taking respective tensor products of the one-dimensional wavelet basis. In contrast, the frequently used higher-dimensional multiresolution basis does not optimally compress the signal in anisotropic smoothness classes. This implies that any coordinatewise thresholded estimator based on such a basis is not able to attain the optimal rate of convergence.

The second part of this paper is devoted to the particular problem of spectral estimation. Throughout the paper we adopt the model of locally stationary time series developed in Dahlhaus (1993). In order to allow least restrictive assumptions on the smoothness of the spectrum we further relax the assumptions of Dahlhaus (1994) to give a definition of the evolutionary spectrum as a function in the L_2-space over the time-frequency plane. Again, our main goal is to define an estimator that adapts to different degrees of smoothness in time and frequency direction, respectively. In contrast to Dahlhaus (1993) and von Sachs and Schneider (1994), who used a local periodogram on segments of length $N = N(T)$ (with $N \to \infty$ as $T \to \infty$ and $N/T \to 0$), here we define a periodogram-like pointwise statistic which can be considered as an empirical version of the local time-dependent spectrum. By this approach we avoid a kind of presmoothing in time direction and get rid of the additional smoothing parameter N, for which a theoretical approach to its optimal choice is still lacking. This overcomes the shortcoming of fixing with N a lower bound for the ratio of the resolution in time and in frequency direction. Instead, to decide which degree of smoothing is appropriate, we project this time-frequency statistic on a suitable wavelet basis and use thresholding of the resulting coefficients. In view of the results in Section 2, in this construction, we use a tensor product basis. The appropriate tuning of the thresholds requires knowledge about the distribution of the empirical coefficients. Using cumulant techniques we prove asymptotic normality in terms of probabilities of large deviations. This implies the asymptotic risk equivalence of monotonic estimators to the case of normally distributed empirical coefficients and suggests the use of thresholding techniques prescribed by existing theory under Gaussian noise.

Finally, to obtain a fully defined threshold rule, it is natural to use some initial estimator of the standard deviation of the empirical coefficients. We show that rather weak assumptions on an initial estimator of the spectral density guarantee near-optimality of the final estimator.

The paper is organized as follows. In Section 2 we derive minimax rates in anisotropic smoothness classes and examine the two mentioned different kinds of multidimensional wavelet bases w.r.t. their appropriateness in such function spaces. In Section 3, after introducing the model of local stationarity and an L_2-generalization of the definition of the evolutionary spectrum, we develop our new estimator and state theorems on rates for its risk. The proofs are contained in Section 4.

2. Optimal estimation in anisotropic smoothness classes

Before we develop a definite estimation method for the spectral density in the next section, we first consider a question of more general importance: we search for a basis that is appropriate for multidimensional estimation problems in situations, where we have possibly different degrees of smoothness in different directions. To do this we consider balls in anisotropic Sobolev spaces and derive minimax rates in a Gaussian white noise model. For simplicity we only consider the two-dimensional case and restrict ourselves to anisotropic Sobolev spaces, although it is obvious that analogous results can be obtained in higher dimensions and for other function classes, like e.g. anisotropic Besov spaces. We show that thresholded wavelet estimators based on a tensor product wavelet basis in $L_2([0,1] \times [0,1])$ attain the optimal rate of convergence, whereas the one-scale multiresolution basis, which is often used in image analysis problems, does not share this property.

Following Nikol'skii (1975), an anisotropic Sobolev space $W^{m_1,m_2}_{p_1,p_2}$ is defined as

$$W^{m_1,m_2}_{p_1,p_2} \;=\; \left\{ f \;\middle|\; \sum_{i=1}^{2} \left(\|f\|_{p_i} \;+\; \|\frac{\partial^{m_i}}{\partial x_i^{m_i}} f\|_{p_i} \right) \;<\; \infty \right\}.$$

In the following we assume that our object of interest f lies in the set

$$\mathcal{F}^{m_1,m_2}_{p_1,p_2} = \mathcal{F}(m_1,m_2,p_1,p_2,C) = \left\{ f \,\middle|\, \sum_{i=1}^{2} \left(\|f\|_{p_i} + \|\frac{\partial^{m_i}}{\partial x_i^{m_i}} f\|_{p_i} \right) \le C \right\} \quad (2.1)$$

for any positive constant C.
Throughout the paper we restrict our considerations to $m_i \ge 1$, $p_i \ge 1$ and $m_i > 1/p_i$, which in particular implies continuity of f.
Since the problem investigated in this section seems to be of general interest in many statistical estimation problems, we do not want to specify any specific observation model. Instead, we assume that function-valued observations $Y(x_1,x_2)$ from the Gaussian white noise model

$$Y(x_1,x_2) = \int_0^{x_2}\int_0^{x_1} f(z_1,z_2)\,dz_1\,dz_2 + \epsilon\, W(x_1,x_2) \quad (2.2)$$

are available. Here W is a Brownian sheet (cf., e.g., Walsh (1986)) and $\epsilon > 0$ is the noise level.

Remark 2.1. In the one-dimensional case it is well-known that the difficulty in estimating f in Gaussian white noise

$$Y(t) = \int_0^t f(s)\,ds + \epsilon W_t, \quad (2.3)$$

where W_t is a standard Wiener process, is closely related to the difficulty in estimating f in non-Gaussian or non-i.i.d. situations, which is actually the interesting problem. Recently, this connection between nonparametric regression and model (2.3) has been established in a decision theoretic manner by Brown and Low (1992). The equivalence between density estimation and some slightly modified version of (2.3) was shown by Nussbaum (1994).
For wavelet estimators this close connection often materializes also at the practical level. So it was shown in Neumann and Spokoiny (1995) for non-Gaussian regression and in Neumann (1994) for spectral density estimation that the empirical coefficients coming from these models are asymptotically normally distributed in a sufficiently strong sense. Then, for certain nonlinear wavelet estimators, it was possible to derive the risk equivalence between model (2.3) and the abovementioned models. We think that the two-dimensional continuous Gaussian model (2.2) will be again an appropriate counterpart for many practically relevant estimation problems.

Assume we have an orthonormal basis of compactly supported wavelets of $L_2[0,1]$, where the functions ϕ and ψ satisfy, for $m \ge \max\{m_1,m_2\}$,

(A1) (i) ϕ and ψ are in C^m,
(ii) $\int \phi(t)\,dt = 1$,
(iii) $\int \psi(t) t^k\,dt = 0$ for $0 \le k \le m-1$.

Such bases are given by Meyer (1991) and Cohen, Daubechies and Vial (1993).
Let V_j be the subspace of $L_2[0,1]$, which is generated by $\{\phi_{jk}\}_k$. It is known that

$$L_2([0,1]\times[0,1]) = \overline{\bigcup_{j=l}^{\infty} V_j \otimes V_j},$$

which shows the possibility to build a basis of $L_2([0,1]\times[0,1])$ from tensor products of functions from a one-dimensional basis $\{\phi_{lk}\}_k \cup \{\psi_{jk}\}_{j\geq l;k}$. Let $W_j = \mathrm{span}\{\psi_{jk}\}_k$. We can write $V_{j^*}^{(2)} = V_{j^*} \otimes\ V_{j^*}$ as

$$\begin{aligned} V_{j^*}^{(2)} &= (V_l \oplus W_l \oplus \cdots \oplus W_{j^*-1}) \otimes (V_l \oplus W_l \oplus \cdots \oplus W_{j^*-1}) \\ &= V_l \otimes V_l \oplus \left(\bigoplus_{j=l}^{j^*-1} (W_j \otimes V_l)\right) \oplus \left(\bigoplus_{j=l}^{j^*-1} (V_l \otimes W_j)\right) \oplus \left(\bigoplus_{j_1,j_2=l}^{j^*-1} (W_{j_1} \otimes W_{j_2})\right) \end{aligned} \tag{2.4}$$

as well as in the form

$$V_{j^*}^{(2)} = V_l \otimes V_l \oplus \bigoplus_{j=l}^{j^*-1} \left[(V_j \otimes W_j) \oplus (W_j \otimes V_j) \oplus (W_j \otimes W_j)\right]. \tag{2.5}$$

According to (2.4) we obtain a basis $\mathcal{B}$ of $L_2([0,1]\times[0,1])$ as

$$\begin{aligned} \mathcal{B} &= \{\phi_{lk_1}(x_1)\phi_{lk_2}(x_2)\}_{k_1,k_2} \cup \left(\bigcup_{j_1\geq l} \{\psi_{j_1k_1}(x_1)\phi_{lk_2}(x_2)\}_{k_1,k_2}\right) \\ &\quad \cup \left(\bigcup_{j_2\geq l} \{\phi_{lk_1}(x_1)\psi_{j_2k_2}(x_2)\}_{k_1,k_2}\right) \cup \left(\bigcup_{j_1,j_2\geq l} \{\psi_{j_1k_1}(x_1)\psi_{j_2k_2}(x_2)\}_{k_1,k_2}\right). \end{aligned} \tag{2.6}$$

Another construction, which corresponds to decomposition (2.5), is given by

$$\begin{aligned} \overline{\mathcal{B}} &= \{\phi_{lk_1}(x_1)\phi_{lk_2}(x_2)\}_{k_1,k_2} \\ &\quad \cup \bigcup_{j\geq l} \{\phi_{jk_1}(x_1)\psi_{jk_2}(x_2), \psi_{jk_1}(x_1)\phi_{jk_2}(x_2), \psi_{jk_1}(x_1)\psi_{jk_2}(x_2)\}_{k_1,k_2}. \end{aligned} \tag{2.7}$$

Note that we can also use different one-dimensional bases to build a two-dimensional basis, which is done in Section 3 in view of the special problem considered there.
It appears that, because of its more appealing structure, basis $\overline{\mathcal{B}}$ is more often used for two-dimensional estimation problems, see, e.g., Delyon and Juditsky (1993), Tribouley (1995) and von Sachs and Schneider (1994). Its use seems to be appropriate in most frequently considered smoothness classes, like e.g. isotropic Sobolev or Besov classes. However, in certain practical problems, for the curve we are interested in we could expect different smoothness properties in different directions. We will show that under such anisotropic smoothness priors basis $\overline{\mathcal{B}}$ is no longer appropriate.
For sake of simplicity we slightly abuse the notation and define $\psi_{l-1,k} := \phi_{lk}$. Further, by μ_I we denote the basis functions in $\mathcal{B}$ using the multiindex $I = (j_1, j_2, k_1, k_2)$. Let $\Theta = \{(\theta_I) \big| \sum_I \theta_I \mu_I \in \mathcal{F}^{m_1,m_2}_{p_1,p_2}\}$. By Parseval's equality we see that the L_2-loss $\|\sum \widehat{\theta}_I\mu_I - f\|^2$ of any estimator $\widehat{f} = \sum \widehat{\theta_I}\mu_I$ in the function space is equal to the l_2-loss $\sum_I(\widehat{\theta}_I - \theta_I)^2$ in the sequence space, where

$$\theta_I = \int\int \mu_I(x_1,x_2)f(x_1,x_2)\,dx_1\,dx_2$$

are the wavelet coefficients of f. We obtain empirical coefficients from the observation model (2.2) as

$$\widetilde{\theta}_I \;=\; \int\int \mu_I(x_1,x_2)\, dY(x_1,x_2) \;=\; \theta_I \,+\, \epsilon\, \xi_I, \tag{2.8}$$

where $\xi_I \sim N(0,1)$ are i.i.d.
First we derive a lower bound for the minimax risk in model (2.2) under the assumption that $f \in \mathcal{F}_{p_1,p_2}^{m_1,m_2}$. Since we are only interested in the optimal *rate*, we can use a simple approach developed in Bretagnolle and Huber (1979). First, we establish the following lemma, which provides a lower bound for the complexity of the set Θ.

Lemma 2.1. *Assume (A1). The set Θ contains a hypercube of sidelength 2ϵ*

$$\Theta_\epsilon \;=\; \{(\theta_I) \,|\, \theta_I \in [-\epsilon,\epsilon] \;\; for\;\; I \in \mathcal{I}_\epsilon \quad and \quad \theta_I = 0 \;\; for \;\; I \notin \mathcal{I}_\epsilon\}$$

with

$$dim(\Theta_\epsilon) \;=\; \#\mathcal{I}_\epsilon \;\asymp\; \epsilon^{-2(m_1+m_2)/(2m_1m_2+m_1+m_2)}.$$

If we now take independent, uniformly distributed priors on $[-\epsilon,\epsilon]$ for $I \in \mathcal{I}_\epsilon$, due to the independence of the $\widetilde{\theta}_I$'s we obtain a Bayes risk of order $\epsilon^{2-2(m_1+m_2)/(2m_1m_2+m_1+m_2)}$. This implies the following theorem.

Theorem 2.1. *Denote by $\widehat{f}$ any estimator of a member $f \in \mathcal{F}_{p_1,p_2}^{m_1,m_2}$. Then*

$$\inf_{\widehat{f}} \sup_{f \in \mathcal{F}_{p_1,p_2}^{m_1,m_2}} \left\{ \mathbb{E}\|\widehat{f} \,-\, f\|^2 \right\} \;\geq\; C\epsilon^{2\vartheta(m_1,m_2)},$$

where

$$\vartheta(m_1,m_2) \;=\; \frac{2m_1m_2}{2m_1m_2 + m_1 + m_2}.$$

To show that this rate is actually attainable, we consider a certain complexity functional $\widetilde{\Omega}_\epsilon$ to be defined further below, which is similar to the modulus of continuity

$$\Omega_\epsilon(\mathcal{B}, \mathcal{F}_{p_1,p_2}^{m_1,m_2}) \;=\; \sup_{f \in \mathcal{F}_{p_1,p_2}^{m_1,m_2}} \left\{ \sum_I \min\{\epsilon^2, \theta_I^2\} \right\} \tag{2.9}$$

considered in Donoho and Johnstone (1994a). There it was shown that Ω_ϵ gives an almost complete information about uniform rates for diagonal estimators in model (2.2).
Two commonly used rules to treat the coefficients are

1) hard thresholding

$$\delta^{(h)}(\widetilde{\theta}_I, \lambda) \;=\; \widetilde{\theta}_I I\left(|\widetilde{\theta}_I| \geq \lambda\right)$$

and

2) soft thresholding

$$\delta^{(s)}(\widetilde{\theta}_I, \lambda) \;=\; \left(|\widetilde{\theta}_I| - \lambda\right)_+ sgn(\widetilde{\theta}_I).$$

In the following $\delta^{(\cdot)}$ is used to (somewhat sloppily) denote either $\delta^{(h)}$ or $\delta^{(s)}$.
Following the developments in Donoho and Johnstone (1994a) we can derive an estimator that attains the rate prescribed by the modulus of continuity $\Omega_\epsilon(\mathcal{B}, \mathcal{F}_{p_1,p_2}^{m_1,m_2})$ up to a factor of $\log(1/\epsilon)$. To prove that the rate $\epsilon^{2\vartheta(m_1,m_2)}$ is exactly attainable, we have to modify Ω_ϵ slightly. First, by Lemma 1 of Donoho and Johnstone (1994a) we can prove that the relation

$$\mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda) \;-\; \theta_I\right)^2 \;\leq\; C\left(\epsilon^2\left(\frac{\lambda}{\epsilon}+1\right)\varphi(\frac{\lambda}{\epsilon}) \;+\; \min\{\lambda^2, \theta_I^2\}\right) \qquad (2.10)$$

holds uniformly in $\lambda \geq 0$ and $\theta_I \in \mathbb{R}$, where φ denotes the standard normal density. This motivates us to define the complexity functional

$$\widetilde{\Omega}_\epsilon(\mathcal{B}, \mathcal{F}_{p_1,p_2}^{m_1,m_2}) \;=\; \inf_{(\lambda_I)} \sup_{f\in\mathcal{F}_{p_1,p_2}^{m_1,m_2}} \left\{\sum_I \left(\epsilon^2\left(\frac{\lambda_I}{\epsilon}+1\right)\varphi(\frac{\lambda_I}{\epsilon}) \;+\; \min\{\lambda_I^2, \theta_I^2\}\right)\right\}. \qquad (2.11)$$

The essential reason why the modulus of continuity Ω_ϵ does not immediately provide an attainable rate for estimators is that it does not take the possible sparsity of the signal into account. In cases, where we have a too large number of potentially important coefficients, we lose an additional log-term as we do not know which are the really important ones. In contrast, the functional $\widetilde{\Omega}_\epsilon$ penalizes such cases of extreme sparsity by the additional terms $(\lambda_I/\epsilon+1)\varphi(\lambda_I/\epsilon)$, which arise from upper estimates of tail probabilities of Gaussian random variables.
The next lemma shows a particular choice of the vector (λ_I), which provides the rate $\epsilon^{2\vartheta(m_1,m_2)}$ for the right-hand side of (2.11).

Lemma 2.2. *Assume (A1). Let $\lambda_{I,\epsilon}$ be such that*

$$\lambda_{I,\epsilon} \;=\; \begin{cases} 0, & \textit{if} \quad j_1 \leq j_1^* \quad \textit{and} \quad j_2 \leq j_2^* \\ \epsilon K_{m_1,m_2}\sqrt{\max\{(j_1-j_1^*)/m_2, (j_2-j_2^*)/m_1\}}, & \textit{otherwise} \end{cases},$$

where

$$2^{j_1^*} \;=\; \epsilon^{-2/(2m_1+1+m_1/m_2)}, \quad 2^{j_2^*} \;=\; \epsilon^{-2/(2m_2+1+m_2/m_1)}$$

and $K_{m_1,m_2} > \sqrt{2(m_1+m_2)\log(2)}$ *is fixed. Then*

$$\sup_{f\in\mathcal{F}_{p_1,p_2}^{m_1,m_2}} \left\{\sum_I \left(\epsilon^2\left(\frac{\lambda_{I,\epsilon}}{\epsilon}+1\right)\varphi(\frac{\lambda_{I,\epsilon}}{\epsilon}) \;+\; \min\{\lambda_{I,\epsilon}^2, \theta_I^2\}\right)\right\} \;=\; O\left(\epsilon^{2\vartheta(m_1,m_2)}\right).$$

Let the $\lambda_{I,\epsilon}$'s be chosen as in Lemma 2.2 and let

$$\widehat{f}_\epsilon \;=\; \sum \delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,\epsilon})\mu_I. \qquad (2.12)$$

Using this Lemma 2.2 in conjunction with (2.10) we can immediately derive the following theorem, which, together with Theorem 2.1, tells us that $\widehat{f}_\epsilon$ is minimax in the class $\mathcal{F}^{m_1,m_2}_{p_1,p_2}$.

Theorem 2.2. *If (A1) is satisfied, then*

$$\sup_{f\in\mathcal{F}^{m_1,m_2}_{p_1,p_2}} \left\{\mathbb{E}\|\widehat{f}_\epsilon - f\|^2\right\} = O\left(\epsilon^{2\vartheta(m_1,m_2)}\right).$$

Although this theorem provides an interesting theoretical result, it turns out to be of limited practical use. The proposed estimator $\widehat{f}_\epsilon$ requires an appropriate tuning of the thresholds $\lambda_{I,\epsilon}$, which strongly depend on the unknown m_1 and m_2. Even if it would be possible to adapt these parameters in our idealized Gaussian white noise model, it is often not obvious how to transfer such a procedure to other noise structures (i.e. with dependencies, non-Gaussianity) which occur in practically relevant estimation problems. One could try to find specific procedures for each particular case, however, it seems to be difficult to find a universal recipe.
An alternative approach that is much less dependent on prior knowledge of m_1 and m_2 is proposed in a series of papers by Donoho and Johnstone, also contained in Donoho *et al.* (1995). First, we analyze the analog of the tail-n-widths (see Donoho *et al.* (1995)) in our two-dimensional function classes.

Lemma 2.3. *Assume (A1). Let $\widetilde{V}_J = \bigoplus_{j_1+j_2=J}(V_{j_1}\otimes V_{j_2})$. Then*

$$\sup_{f\in\mathcal{F}^{m_1,m_2}_{p_1,p_2}} \left\{\|f - Proj_{\widetilde{V}_J} f\|^2\right\} = O\left(2^{-J\gamma(m_1,m_2,p_1,p_2)}\right),$$

where $\gamma(m_1,m_2,p_1,p_2) = \frac{2m_1m_2+m_1+m_2-2m_1/\widetilde{p}_2-2m_2/\widetilde{p}_1}{m_1+m_2}$, $\quad \widetilde{p}_i = \min\{p_i, 2\}$.

If we now choose J_ϵ sufficiently large, we are able to obtain

$$\sum_{I:\, j_1+j_2>J_\epsilon} \theta_I^2 = O\left(\epsilon^{2\vartheta(m_1,m_2)}\right), \tag{2.13}$$

i.e., the truncation of the wavelet series does not affect the desired rate of the estimator. Define $\mathcal{K}_\epsilon = \{I=(j_1,j_2,k_1,k_2) \mid j_1+j_2 \le J_\epsilon\}$. We consider the estimator

$$\widehat{\widehat{f}}_\epsilon = \sum_{I\in\mathcal{K}_\epsilon} \delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_\epsilon)\mu_I, \tag{2.14}$$

where

$$\lambda_\epsilon = \epsilon\sqrt{2\log(\#\mathcal{K}_\epsilon)}.$$

Using Lemma 2.2, Lemma 2.3 and (2.10) we obtain the following theorem.

Theorem 2.3. *Assume (A1) and $2^{J_\epsilon} = O(\epsilon^{-\eta})$ for any $\eta < \infty$. Then*

$$\sup_{f\in\mathcal{F}^{m_1,m_2}_{p_1,p_2}} \left\{\mathbb{E}\|\widehat{\widehat{f}}_\epsilon - f\|^2\right\} = O\left((\epsilon^2\log(1/\epsilon))^{\vartheta(m_1,m_2)}\right)$$

over all (m_1,m_2,p_1,p_2) satisfying $2^{-J_\epsilon\gamma(m_1,m_2,p_1,p_2)} \le \epsilon^{2\vartheta(m_1,m_2)}$.

Hence, the estimator $\widehat{\widehat{f}}_\epsilon$ is minimax up to a factor of $\log(1/\epsilon)$ over a wide range of function classes.

In the rest of this section we will briefly examine the basis $\overline{\mathcal{B}}$ w.r.t. their capability of data compression in anisotropic Sobolev spaces. The following lemma states a result on the decay of the modulus of continuity Ω_ϵ for this basis.

Lemma 2.4. *It holds that*

$$\Omega_\epsilon(\overline{\mathcal{B}}, \mathcal{F}_{p_1,p_2}^{m_1,m_2}) \asymp \epsilon^{2\overline{\vartheta}(m_1,m_2)},$$

where

$$\overline{\vartheta}(m_1,m_2) \;=\; \min\{\frac{m_1}{m_1+1}, \frac{m_2}{m_2+1}\}.$$

It can be easily shown that $\overline{\vartheta}(m_1,m_2) = \vartheta(m_1,m_2)$ if $m_1 = m_2$ and $\overline{\vartheta}(m_1,m_2) < \vartheta(m_1,m_2)$ if $m_1 \neq m_2$. The rate $\overline{\vartheta}(m_1,m_2)$ is the usual one for a two-dimensional estimation problem in isotropic smoothness classes with degree of smoothness $\overline{m} = \min\{m_1,m_2\}$.
We have already seen that basis $\mathcal{B}$ provides an optimal data compression in the sense that $\tilde{\Omega}_\epsilon(\mathcal{B}, \mathcal{F}_{p_1,p_2}^{m_1,m_2})$ decays at the same rate as the minimax risk in $\mathcal{F}_{p_1,p_2}^{m_1,m_2}$. To make a comparison between the two bases in statistical terms we restrict our consideration to thresholded diagonal estimators in both cases. Let $\overline{\mathcal{B}} = \{\overline{\mu}_I\}$ and let $\overline{\theta}_I$ and $\tilde{\overline{\theta}}_I$ denote the corresponding true and empirical coefficients, respectively. By simple calculations we can show that

$$\inf_\lambda \left\{ \mathbb{E}\left(\delta^{(\cdot)}(\tilde{\overline{\theta}}_I, \lambda) \;-\; \overline{\theta}_I \right)^2 \right\} \;\geq\; C \min\{\epsilon^2, \overline{\theta}_I\}. \tag{2.15}$$

Hence, we will get a lower bound for the risk of thresholded diagonal estimators simply by observing the rate of decay of Ω_ϵ. The following theorem is an immediate consequence of Lemma 2.4 and (2.15).

Theorem 2.4. *Assume (A1). Then*

$$\sup_{f\in\mathcal{F}_{p_1,p_2}^{m_1,m_2}} \inf_{(\lambda_I)} \left\{ \mathbb{E}\| \sum \delta^{(\cdot)}(\tilde{\overline{\theta}}_I, \lambda_I)\overline{\mu}_I \;-\; f\|^2 \right\} \;\geq\; C\epsilon^{2\overline{\vartheta}(m_1,m_2)}.$$

Hence, we get that diagonal estimators based on basis $\overline{\mathcal{B}}$ are never better than those based on $\mathcal{B}$, and they are worse if $m_1 \neq m_2$. At this point we want to remark that there exists an attempt to construct higher-dimensional multiresolution bases for anisotropic smoothness classes. Berkolajko and Novikov (1992) obtained such a basis by properly connecting levels j_1 and j_2 in dependence on the relation between m_1 and m_2. However, as this approach depends strongly on the latter relation, it does not provide a *universal* basis which is optimal for a greater range of smoothness classes. The adaptive choice of an appropriate basis, which in principle seems to be possible in view of results by Donoho and Johnstone (1994b), would call for another step in the estimation process.

3. Adaptive Estimation of Evolutionary Spectra

To address the problem of adaptively estimating the time-dependent spectrum of a non-stationary time series, we start with citing the definition of a locally stationary process, as given in Dahlhaus (1993). Note that this generalizes the Cramér representation of a stationary stochastic process (see Priestley, 1981, e.g.).

Definition 3.1. A sequence of stochastic processes $X_{t,T}\,(t=1,\dots,T)$ is called *locally stationary* if there exists a representation

$$X_{t,T} = \mu(\frac{t}{T}) + \int_{-\pi}^{\pi} A(\frac{t}{T},\omega)\,\exp(i\omega t)\,d\xi(\omega)\,, \tag{3.1}$$

where

(i) $\xi(\omega)$ is a stochastic process on $[-\pi,\pi]$ with $\overline{\xi(\omega)} = \xi(-\omega)$, $\mathbb{E}\xi(\omega) = 0$ and orthonormal increments, i.e. $cov(d\xi(\omega),\ d\xi(\omega')) = \delta(\omega-\omega')d\omega$, $cum\{d\xi(\omega_1),\dots,d\xi(\omega_k)\} = \eta(\sum_{j=1}^{k}\omega_j)\ h_k(\omega_1,\dots,\omega_{k-1})\ d\omega_1\ \dots\ d\omega_k$, where $cum\{\dots\}$ denotes the cumulant of oder k, $|h_k(\omega_1,\dots,\omega_{k-1})| \le const_k$ for all k (with $h_1=0, h_2(\omega)=1$) and $\eta(\omega) = \sum_{j=-\infty}^{\infty}\delta(\omega+2\pi j)$ is the period 2π extension of the Dirac delta function.

(ii) $A(u,\omega)$ is a function on $[0,1]\times[-\pi,\pi]$ which is 2π-periodic in ω, with $A(u,-\omega) = \overline{A}(u,\omega)$.

Remark 3.1. In Dahlhaus (1993) a slightly more general definition of a locally stationary process was given. There, the representation in (3.1) is based on a sequence of functions $A^o_{t,T}(\omega)$ instead of the function $A(u,\omega)$, the difference of which has to fulfill: $\sup_{t,\omega}|A^o_{t,T}(\omega) - A(t/T,\omega)| \le K\,T^{-1}$, for some positive constant K.
Note that with this, the class of autoregressive processes with time-varying coefficients now is included in the class of locally stationary processes.
In our work, for reasons of notational convenience, we do not want to adopt this more general definition, noting that all results will continue to hold for the broader class.

Note that, as in Dahlhaus (1993) and von Sachs and Schneider (1994), for simplicity we assume that $\mu(u)=0$, i.e. we do not treat the problem of estimating the mean of the time series. In comparison to Dahlhaus (1993) and (1994), here, our smoothness assumptions on $A(u,\omega)$ are slightly relaxed: Basically we like to impose minimal smoothness as being of bounded variation on $U\times\Pi := [0,1]\times[-\pi,\pi]$ (which is made precise in Assumption (A2)). For technical reasons, in order to facilitate proofs, we impose an additional smoothness condition on the decay of the Fourier coefficients of $A(u,\omega)$ as a function of ω, which implies continuity of A in ω.

Before proceeding with the introduction of both evolutionary spectrum of $\{X_{t,T}\}$ and a suitable fully adaptive spectral estimate, we gather the assumptions that are necessary to end up with a more general definition of the spectrum and for deriving our asymptotic results:

Definition 3.2. (Total variation on $U \times \Pi := [0,1] \times [-\pi, \pi]$):

$$TV_{U\times\Pi}(f) := \sup \sum_i \sum_j |f(u_i, \omega_j) - f(u_i, \omega_{j-1}) - f(u_{i-1}, \omega_j) + f(u_{i-1}, \omega_{j-1})|\,,$$

where the supremum is to be taken over all partitions of $U \times \Pi$.

Now we impose the following assumptions:

(A2) a) $A(u,\omega)$ has bounded total variation on $U \times \Pi$, i.e. $TV_{U\times\Pi}(A) < \infty$.
b) $\sup_u TV_{[-\pi,\pi]}(A(u,.)) < \infty$ and $\sup_\omega TV_{[0,1]}(A(.,\omega)) < \infty$.
c) $\sup_{u,\omega} |A(u,\omega)| < \infty$.
d) $\inf_{u,\omega} |A(u,\omega)| \geq \kappa$ for some $\kappa > 0$.

(A3) Let $\widehat{A}(u,s) := 1/(2\pi) \int A(u,\omega) \exp(i\omega s)\, d\omega$, $s \in \mathbb{Z}, u \in [0,1]$.
Then: $\sup_u \sum_s |\widehat{A}(u,s)| < \infty$.

(A4) a) $\phi(u)$, $\psi(u)$, $\widetilde{\phi}(\omega)$ and $\widetilde{\psi}(\omega)$ have bounded total variation on $[0,1]$ and $[-\pi,\pi]$, respectively.
b) Further, $\sum_s |\widehat{\widetilde{\phi}}(s)| < \infty$ and $\sum_s |\widehat{\widetilde{\psi}}(s)| < \infty$.

(A5) $\sup_{1\leq t_1 \leq T} \left\{ \sum_{t_2,\ldots,t_k=1}^T |cum(X_{t_1,T}, \ldots, X_{t_k,T})| \right\} \leq C^k (k!)^{1+\gamma}$ for all $k = 2, 3, \ldots,$ where $\gamma \geq 0$.

Note that these are somewhat minimal conditions part of which might be fulfilled simply by restricting A to be member of the specific smoothness class under consideration (anisotropic Sobolev, Hölder,...). In our case of Sobolev restrictions (A2) (b) and (c) and (A3) are implications of the considered Sobolev smoothness, so are (A4) (a) and (b) a consequence of (A1), with $m \geq \max\{m_1, m_2\} \geq 1$.
We like to mention that this minimal smoothness of A is sufficient to ensure the locally stationary behavior of the process, in the sense that we end up with a spectrum which is uniquely defined in some L^2- rather than in an almost everywhere sense. However, for reasons of completeness, we like to also give this stronger definition of the evolutionary spectrum which, under the appropriate stronger smoothness of A, was considered by Dahlhaus (1993):

Definition 3.3. As *evolutionary spectrum* of $\{X_{t,T}\}$ given in (3.1) we define for $u \in (0,1)$

$$f(u,\omega) = \lim_{T\to\infty} \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} cov\{X_{[uT-s/2],T}; X_{[uT+s/2],T}\} \exp(-i\omega s),$$

where the $X_{t,T}$'s are given by (3.1) with $A(t/T,\omega) = A(0,\omega)$ for $t < 1$ and $A(t/T,\omega) = A(1,\omega)$ for $t > T$.

By Dahlhaus (1993), Theorem 2.2, if $A(u,\omega)$ is differentiable in u and ω (with uniformly bounded derivatives), then

$$f(u,\omega) = |A(u,\omega)|^2, \; u \in (0,1) \qquad \text{a.e. in } \omega. \tag{3.2}$$

Whenever this condition on A is fulfilled we shall understand the given limit in (3.2) as pointwise in u and ω.
More generally, however, we like to show that, if we turn to the L_2-limit, equation (3.2) still holds, in the $L_2(du, d\omega)$-sense on $U \times \Pi$:

Theorem 3.1. *Under assumptions (A2) and (A3),*

$$\lim_{T\to\infty} \int_0^1 \int_{-\pi}^{\pi} \{ \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} [cov\{X_{[uT-s/2],T}; X_{[uT+s/2],T}\} \exp(-i\omega s)] - |A(u,\omega)|^2 \}^2 \, d\omega \; du = 0.$$

An intermediate result, finally, which is in the $L^2(d\omega)$-sense, but pointwise in $u \in (0,1)$, is given by Dahlhaus (1994), Theorem 2.2, where uniform Lipschitz-continuity of $A(u,\omega)$ in both components with Lipschitz exponent $\alpha > 1/2$ is needed.

For the particular context of our work, we now restrict to the anisotropic Sobolev class as introduced in Section 2, i.e. we assume that f is a member of this class by assuming that $A(u,\omega)$ is:

$A \in \mathcal{F}^{m_1,m_2}_{p_1,p_2}(C)$ with $m_i \geq 1, p_i \geq 1$ and $m_i > 1/p_i$.

We note that with this f is in any $L_p(U \times \Pi)$ - space (due to the continuity in each argument), i.e. in particular in L_2.

Now we turn to the problem of estimating the evolutionary spectrum f.

The first step in our inference about f is to transfer the information $\{X_{1,T}, \ldots, X_{T,T}\}$ given in the time domain to the time-frequency domain. One possibility, as chosen by Dahlhaus (1993) and also in von Sachs and Schneider (1994), is to consider a localized periodogram, localized by introducing segments of length $N = N(T)$, where $N \to \infty$ as $T \to \infty$ but $N/T \to 0$. One problem with this approach is that the segment length N is an additional parameter, whose optimal choice depends on the relation between the smoothness in time and frequency direction. Here we intend to develop a fully adaptive approach: By wavelet thresholding the procedure should be able to automatically adapt to the right degree of resolution in both time and frequency direction. Note that these are, of course, not independent, but stand in a reciprocal relationship due to the uncertainty principle: the more accurate we try to estimate $f(u,\omega)$ in time direction, the less accurate can we estimate it in frequency direction and vice versa, cf. Priestley (1981, p. 835).
To this end, by a straightforward analogy to the definition of the spectral density we introduce a periodogram-like statistic $I_{t,T}, 1 \leq t \leq T$, which is different to the localized periodogram of von Sachs and Schneider (1994):

$$I_{t,T}(\omega) = \frac{1}{2\pi} \sum_{|s/2| \leq \min\{t-1, T-t\}} X_{[t-s/2],T} X_{[t+s/2],T} \exp(-i\omega s). \qquad (3.3)$$

Note that $I_{t,T}$ can be considered as a preliminary "estimate" which is even more fluctuating than the classical periodogram is.

In von Sachs and Schneider (1994) part of the localization was delivered by summation over certain time points in segments of chosen length N before the actual local smoothing was performed by wavelet thresholding. Thus, inherently a lower bound was fixed for the resolution in time which obviously had consequences also for the performance in frequency direction: The larger N the worse is the time resolution, but the better can low-frequency components be detected, and vice versa. Here, in our new approach, we avoid a two-fold smoothing: projection of these "rough periodograms" $I_{t,T}$ on an appropriate wavelet basis will do the whole task of adaptive local smoothing!

To give the link to the previous section on anisotropic smoothness classes, with this particular task, we are confronted with a two-dimensional estimation problem, where the axes have a special meaning, time and frequency, respectively. It seems reasonable to design the estimation method in such a way that it takes different degrees of smoothness into these two directions into account.

As we have seen in the preceding section, we obtain an appropriate wavelet basis according to the definition of basis $\mathcal{B} = \{\mu_I(u,\omega)\}$. We get such a basis as tensor product of two bases, where in time direction we choose a wavelet basis on the interval $\{\phi_{lk}\}_k \cup \{\psi_{jk}\}_{j\geq l,k}$ (e.g. boundary-corrected Meyer wavelets, see Meyer (1991), or those of Cohen, Daubechies, Vial (1993)). In frequency direction a periodic basis $\{\widetilde{\phi}_{lk}\}_k \cup \{\widetilde{\psi}_{jk}\}_{j\geq l,k}$ is used (as proposed in Daubechies (1992, Chapter 9.3)). As an example, we like to mention the orthogonal periodized Battle-Lemarié spline wavelets (as in von Sachs and Schneider (1994)), though these have "numerical compact support", only, but our proofs will only slightly change with these. For notational convenience we write again $\psi_{l-1,k}$ and $\widetilde{\psi}_{l-1,k}$ for ϕ_{lk} and $\widetilde{\phi}_{lk}$, respectively. Whenever it is not misleading, we use the multiindex $I = (j_1, j_2, k_1, k_2)$.

In addition to the "true" wavelet coefficients θ_I of $f(u,\omega)$

$$\theta_I \;=\; \int_{U\times\Pi} f(u,\omega)\,\mu_I(u,\omega)\,du\;d\omega \;=\; \int_{U\times\Pi} f(u,\omega)\,\psi_{j_1k_1}(u)\,\widetilde{\psi}_{j_2k_2}(\omega)\,du\;d\omega\,, \tag{3.4}$$

we define empirical wavelet coefficients as follows:

$$\widetilde{\theta}_I \;=\; \sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} \psi_{j_1k_1}(u)\,du \int_{-\pi}^{\pi} \widetilde{\psi}_{j_2k_2}(\omega) I_{t,T}(\omega)\,d\omega. \tag{3.5}$$

In the special case of a stationary time series, the advantage of the tensor product basis over the multiresolution basis becomes apparent. Then all coefficients θ_I with $j_1 \neq l-1$ are equal to zero, whereas $\theta_{(l-1,k_1,j_2,k_2)} \asymp 2^{-l/2}\theta_{j_2k_2}$, where $\theta_{j_2k_2} = \int f(\omega)\widetilde{\psi}_{j_2k_2}(\omega)\,d\omega$ are the wavelet coefficients of the (one-dimensional) spectral density $f(\omega) = f(u,\omega)$. In view of the results from Section 2 it is obvious that in estimating $f(u,\omega)$, which is constant in u, we can obtain the same rate as in Neumann (1994) in the stationary case.

In the following we intend to derive asymptotic normality of the empirical coefficients by the method of cumulants. It turns out that a simple central limit theorem would not be sufficient for proving risk equivalence between our thresholded wavelet estimator and the case of Gaussian noise. In view of quite a large number of coefficients which cannot be a priori neglected in cases of "inhomogeneous smoothness", we have to choose the threshold somewhat higher than the noise level, i.e. of larger order than the standard deviation of the empirical coefficients. Accordingly, we need some formulation of asymptotic normality, which puts special emphasis on moderate and large deviations.

Let σ_I^2 denote the variance of $\widetilde{\theta}_I$. In contrast to a central limit theorem, where it would be sufficient to show that $cum_n(\widetilde{\theta}_I/\sigma_I) = o(1)$ holds for each particular $n \geq 3$, here we need a stronger estimate for the higher order cumulants. For the reader's convenience we quote a lemma from Neumann (1994), which provides appropriate estimates for general quadratic forms.

Lemma 3.1. *Let*

$$\eta_T \;=\; \underline{X}' M \underline{X},$$

where

$$\underline{X} \;=\; (X_{1,T}, \ldots, X_{T,T})', \quad M \;=\; ((M_{st}))_{s,t=1,\ldots,T}, \quad M_{st} \;=\; M_{ts}.$$

Further, let

$$\xi_T \;=\; \underline{Y}' M \underline{Y},$$

where $\underline{Y} = (Y_1, \ldots, Y_T)' \sim N(0, Cov(\underline{X}))$.
Then, under (A5),

$$cum_n(\eta_T) \;=\; cum_n(\xi_T) \;+\; R_n$$

holds for $n \geq 2$, *where*

$$\text{(i)} \quad |cum_n(\xi_T)| \;\leq\; var(\xi_T) 2^{n-2}(n-1)! \left[\lambda_{\max}\left(M\, Cov(\underline{X})\right)\right]^{n-2}$$

$$\text{(ii)} \quad R_n \;\leq\; 2^{n-2} C^{2n}((2n)!)^{1+\gamma} \max_{s,t}\{|M_{st}|\}\, \widetilde{M}\; \|M\|_\infty^{n-2},$$

$$\widetilde{M} \;=\; \sum_s \max_t\{|M_{st}|\}, \quad \|M\|_\infty \;=\; \max_s \left\{\sum_t |M_{st}|\right\}.$$

In the following we are able to show asymptotic normality for all coefficients $\widetilde{\theta}_I$ with $2^{j_1+j_2} = o(T)$ and $j_2^{-1} = o(1)$.
Fix some $\delta > 0$. We define

$$\mathcal{I}_T \;=\; \left\{ I \,\middle|\, 2^{j_1+j_2} \leq T^{1-\delta} \right\}. \tag{3.6}$$

Making use of Lemma 3.1 we obtain the following result for the empirical coefficients.

Lemma 3.2. *Assume (A1) through (A5). Then*

$$\text{(i)} \quad \mathbb{E}\,\widetilde{\theta}_I \;=\; \theta_I \;+\; o(T^{-1/2}),$$

$$\text{(ii)} \quad var(\widetilde{\theta}_I) \;=\; 2\pi\; T^{-1} \int_{U\times\Pi} \{f(u,\omega)\,\psi_{j_1k_1}(u)\}^2\, du\; \widetilde{\psi}_{j_2k_2}(\omega)\; [\widetilde{\psi}_{j_2k_2}(\omega) \;+\; \widetilde{\psi}_{j_2k_2}(-\omega)]\, d\omega \;+\; o(T^{-1}) \;+\; O(2^{-j_2}T^{-1}),$$

$$\text{(iii)} \quad \left|cum_n(\widetilde{\theta}_I)\right| \;\leq\; (n!)^{2+2\gamma} C^n T^{-1} (T^{-1} 2^{(j_1+j_2)/2} \log(T))^{n-2} \qquad \textit{for } n \geq 3$$

and appropriate $C \;>\; 0$ *uniformly in* $I \in \mathcal{I}_T$.

Using Lemma 1 in Rudzkis, Saulis and Statulevicius (1978) we now obtain the desired version of asymptotic normality.

Proposition 3.1. *Assume (A1) through (A5). Let* $\Delta_T = (\log T)^\lambda$ *for any fixed* $0 < \lambda < \infty$. *Then*

$$P\left(\pm(\widetilde{\theta}_I - \theta_I)/\sigma_I \geq x\right) = (1 - \Phi(x))(1 + o(1))$$

holds uniformly in $-\infty \leq x \leq \Delta_T$ *and* $I \in \mathcal{I}_T \cap \{I \mid 2^{j_2} \geq T^\rho\}$ *for* $\rho > 0$ *arbitrarily small, where* $\Phi(x) = \int_{-\infty}^x \varphi(t)\,dt$ *denotes the standard normal cumulative distribution function.*

Let

$$\mathcal{I}_T^0 = \{I \in \mathcal{I}_T \mid (j_1, j_2) \neq (l-1, l-1)\}.$$

We consider the estimator

$$\widehat{f}(u,\omega) = \sum_{I \in \mathcal{I}_T^0} \delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,T})\mu_I(u,\omega),$$

where the thresholds $\lambda_{I,T}$ are specified below. As usually done, we do not shrink the coefficients from the coarsest level $(j_1, j_2) = (l-1, l-1)$. This seems to be reasonable in view of our assumption (A2), d), which implies that the spectrum is bounded away from zero.
In order to establish the equivalence to the case of Gaussian noise, we consider the following approximating model for our empirical coefficients:

$$\widetilde{\xi}_I = \theta_I + \sigma_I \varepsilon_I, \quad I \in \mathcal{I}_T,$$

where $\varepsilon_I \sim N(0,1)$.
Essentially by integration by parts, due to Proposition 3.1 we obtain the following assertion.

Proposition 3.2. *Assume (A1) through (A5). Then, for arbitrary nonrandom thresholds* $\lambda_{I,T} = O(T^{-1/2}\sqrt{\log(T)})$,

$$\sum_{I \in \mathcal{I}_T} \mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,T}) - \theta_I\right)^2 = (1 + o(1)) \sum_{I \in \mathcal{I}_T} \mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\xi}_I, \lambda_{I,T}) - \theta_I\right)^2 + O(T^{-\vartheta(m_1,m_2)}).$$

This asymptotic risk equivalence enables us to derive the following theorem. Recall that $\vartheta(m_1, m_2)$ was defined in Theorem 2.1 and $\gamma(m_1, m_2, p_1, p_2)$ in Lemma 2.3, respectively.

Theorem 3.2. *Assume (A1) through (A5) and* $(1-\delta)\,\gamma(m_1, m_2, p_1, p_2) \geq \vartheta(m_1, m_2)$ *with* δ *as in (3.6).*
Further, assume that, for some $\gamma_T \to 1$,

$$\gamma_T \sigma_I \sqrt{2\log(\#\mathcal{I}_T^0)} \leq \lambda_{I,T} \leq C T^{-1/2}\sqrt{\log(T)}$$

holds for $I \in \mathcal{I}_T^*$, $\mathcal{I}_T^* \subseteq \mathcal{I}_T^0$, *where* $\#(\mathcal{I}_T^0 \setminus \mathcal{I}_T^*) = O(T^{1-\vartheta(m_1,m_2)})$. *Then*

$$\mathbb{E}\,\|\widehat{f} - f\|^2_{L_2([0,1]\times[-\pi,\pi])} = O\left((\log(T)/T)^{-\vartheta(m_1,m_2)}\right).$$

There are many possibilities for m_1, m_2, p_1 and p_2 to fulfill $\gamma(m_1,m_2,p_1,p_2) > \vartheta(m_1,m_2)$, for example, if $m_i \geq 2/p_i$. Then we can find some sufficiently small $\delta > 0$, such that the assumption of Theorem 3.2 is satisfied. Hence, our estimator is simultaneously nearly optimal over a wide range of smoothness classes.
Although Theorem 3.2 is of certain theoretical interest, it is not very helpful for practical purposes, because the definition of the estimator $\widehat{f}$ depends on the unknown quantities σ_I. It is a natural idea to use some initial estimates of them to construct a fully adaptive procedure. Let $\widehat{\lambda}_I$ be any random thresholds and $\widehat{\widehat{f}}$ be the same estimator as $\widehat{f}$ with these random thresholds. The next theorem characterizes the performance of such an estimator under a weak assumption on the random thresholds.

Theorem 3.3. *Assume (A1) through (A5). Let* $(1-\delta)\,\gamma(m_1,m_2,p_1,p_2) \geq \vartheta(m_1,m_2)$. *Assume that, for some* $\gamma_T \to 1$,

$$\sum_{I \in \mathcal{I}_T^0} \mathbb{E}(\widetilde{\theta}_I^2 \,+\, 1) I\left(\widehat{\lambda}_I \notin [\gamma_T \sigma_I \sqrt{2\log(\#\mathcal{I}_T^0)}, C T^{-1/2}\sqrt{\log(T)}]\right) \;=\; O(T^{-\vartheta(m_1,m_2)}). \tag{3.7}$$

Then

$$\mathbb{E}\|\widehat{\widehat{f}} \,-\, f\|^2_{L_2([0,1]\times[-\pi,\pi])} \;=\; O\left((\log(T)/T)^{-\vartheta(m_1,m_2)}\right).$$

Remark 3.2.
(i) By Cauchy-Schwarz, (3.7) is obviously satisfied, if

$$P\left(\widehat{\lambda}_I \notin [\gamma_T \sigma_I \sqrt{2\log(\#\mathcal{I}_T^0)}, C T^{-1/2}\sqrt{\log(T)}]\right) \;=\; O(T^{-4}).$$

holds for $I \in \mathcal{I}_T^*$, where $\#(\mathcal{I}_T^0 \setminus \mathcal{I}_T^*) = O(T^{1-\vartheta(m_1,m_2)})$.
(ii) If the assumptions of the Theorems 3.2 and 3.3 are to hold uniformly, then all assertions will hold uniformly in the class $\mathcal{F}_{p_1,p_2}^{m_1,m_2}$.

To end up with a fully automatic estimator, we still have to find a practicable rule for the thresholds λ_I. All we need are asymptotic majorants of $\sigma_I\sqrt{2\log(\#\mathcal{I}_T^0)}$, which are also of order $T^{-1/2}\sqrt{\log(T)}$. This can be achieved by plugging in some consistent preliminary estimate $\widetilde{f}$ into the asymptotic formula for the variance of the empirical coefficients, which is given in Lemma 3.2. Then we can use the thresholds $\widehat{\lambda}_I = \widehat{\sigma}_I\sqrt{2\log(\#\mathcal{I}_T^0)}$, with $\widehat{\sigma}_I$ as in Lemma 3.2(ii). It turns out that (3.7) will be satisfied under weak assumptions on the time series and the estimator $\widetilde{f}$.

4. Proofs

Proof of Lemma 2.1. Let j_1^* and j_2^* be chosen such that

$$\begin{aligned} 2^{j_1^*} &\le C_0\epsilon^{-2m_2/(2m_1m_2+m_1+m_2)} < 2^{j_1^*+1}, \\ 2^{j_2^*} &\le C_0\epsilon^{-2m_1/(2m_1m_2+m_1+m_2)} < 2^{j_2^*+1} \end{aligned}$$

hold for some C_0 chosen at the end of this proof. Define

$$\mathcal{I}_\epsilon = \{I = (j_1, j_2, k_1, k_2) \mid (j_1, j_2) = (j_1^*, j_2^*)\}.$$

It is obvious that $\mathcal{I}_\epsilon$ satisfies

$$\#\mathcal{I}_\epsilon \asymp 2^{j_1^*+j_2^*} \asymp \epsilon^{-2(m_1+m_2)/(2m_1m_2+m_1+m_2)}.$$

It remains to show that, for an appropriate choice of C_0, the relation $\Theta_\epsilon \subseteq \Theta$ holds. Let $f = \sum \theta_I \mu_I$ be arbitrary with $(\theta_I) \in \Theta_\epsilon$. Then we obtain

$$\|f\|_{p_i} \le \|f\|_\infty \le C\epsilon 2^{(j_1^*+j_2^*)/2} \le CC_0\epsilon^{2m_1m_2/(2m_1m_2+m_1+m_2)} \tag{4.1}$$

and

$$\|\frac{\partial^{m_i}}{\partial x_i^{m_i}} f\|_{p_i} \le \|\frac{\partial^{m_i}}{\partial x_i^{m_i}} f\|_\infty \le C\epsilon 2^{j_i^* m_i} 2^{(j_1^*+j_2^*)/2} \le CC_0. \tag{4.2}$$

For C_0 small enough we obtain $f \in \mathcal{F}_{p_1,p_2}^{m_1,m_2}$, which implies $\Theta_\epsilon \subseteq \Theta$. $\square$

Proof of Lemma 2.2. Let $\mathcal{I}_\epsilon$ be chosen as in the proof of Lemma 2.1 and let

$$\theta_I^* = \begin{cases} \epsilon, & \text{if } I \in \mathcal{I}_\epsilon \\ 0 & \text{otherwise} \end{cases}.$$

We have seen in the proof of Lemma 2.1 that $(\theta_I^*) \in \Theta$ holds, which implies

$$\Omega_\epsilon(\mathcal{B}, \mathcal{F}_{p_1,p_2}^{m_1,m_2}) \ge \epsilon^2 \#\mathcal{I}_\epsilon \ge C\epsilon^{2\vartheta(m_1,m_2)}.$$

Since $\Omega_\epsilon(\mathcal{B}, \mathcal{F}_{p_1,p_2}^{m_1,m_2}) \le \widetilde{\Omega}_\epsilon(\mathcal{B}, \mathcal{F}_{p_1,p_2}^{m_1,m_2})$, we have a lower bound for $\widetilde{\Omega}_\epsilon(\mathcal{B}, \mathcal{F}_{p_1,p_2}^{m_1,m_2})$.
Let now $f \in \mathcal{F}_{p_1,p_2}^{m_1,m_2}$ be arbitrary.
Let $j_1 \ge l$ and $x_{(j_1,k_1)} \in supp(\psi_{j_1k_1})$. Then, by Taylor's formula,

$$\begin{aligned} &\int \psi_{j_1k_1}(x_1) f(x_1, x_2)\, dx_1 \\ &= \int \psi_{j_1k_1}(x_1) \left[\int_{x_{(j_1,k_1)}}^{x_1} \frac{(x_1 - z)^{m_1-1}}{(m_1-1)!} \frac{\partial^{m_1}}{\partial x_1^{m_1}} f(z, x_2)\, dz\right] dx_1 \\ &= O\left(2^{-j_1(m_1-1)}\right) \int |\psi_{j_1k_1}(x_1)|\, dx_1 \int_{supp(\psi_{j_1k_1})} \left|\frac{\partial^{m_1}}{\partial x_1^{m_1}} f(z, x_2)\right| dz \\ &= O\left(2^{-j_1(m_1-1/2)}\right) \int_{supp(\psi_{j_1k_1})} \left|\frac{\partial^{m_1}}{\partial x_1^{m_1}} f(z, x_2)\right| dz, \end{aligned}$$

which implies, since every basis function μ_I overlaps only with a finite number of basis functions from the same scale (j_1, j_2), that

$$
\begin{aligned}
&\sum_{k_1,k_2} \left|\theta_{(j_1 j_2 k_1 k_2)}\right|^p \\
&= \sum_{k_1,k_2} \left| \int \psi_{j_2 k_2}(x_2) \int \psi_{j_1 k_1}(x_1) f(x_1, x_2)\, dx_1\, dx_2 \right|^p \\
&\leq C \sum_{k_1,k_2} 2^{-j_1(m_1-1/2)p} 2^{j_2 p/2} \left[\iint_{supp(\mu_{(j_1 j_2 k_1 k_2)})} \left| \frac{\partial^{m_1}}{\partial x_1^{m_1}} f(x_1, x_2) \right| dx_1\, dx_2 \right]^p \\
&\leq C \sum_{k_1,k_2} 2^{-j_1 m_1 p} 2^{(j_1+j_2)p/2} \iint_{supp(\mu_{(j_1 j_2 k_1 k_2)})} \left| \frac{\partial^{m_1}}{\partial x_1^{m_1}} f(x_1, x_2) \right|^p dx_1\, dx_2\, * \\
&\qquad * \left(mes(supp(\mu_{(j_1 j_2 k_1 k_2)})) \right)^{p-1} \\
&\leq C 2^{-j_1 m_1 p} 2^{(j_1+j_2)(1-p/2)} \left\| \frac{\partial^{m_1}}{\partial x_1^{m_1}} f \right\|_{p_1}^p \\
&= O\left(2^{-j_1 m_1 p} 2^{(j_1+j_2)(1-p/2)}\right) \qquad (4.3)
\end{aligned}
$$

for $p \leq p_1$. By analogous calculations we can show that

$$
\sum_{k_1,k_2} \left|\theta_{(j_1 j_2 k_1 k_2)}\right|^p = O\left(2^{-j_2 m_2 p} 2^{(j_1+j_2)(1-p/2)}\right) \qquad (4.4)
$$

holds for $j_2 \geq l, \quad p \leq p_2$.
Let j_1^* and j_2^* be such that $2^{j_1^*} = \epsilon^{-2/(2m_1+1+m_1/m_2)}$ and $2^{j_2^*} = \epsilon^{-2/(2m_2+1+m_2/m_1)}$. We decompose the set $\mathcal{J} = \{(j_1, j_2) \mid j_1 \geq l, j_2 \geq l\}$ into the following three sets:

$$
\begin{aligned}
\mathcal{J}_1 &= \{(j_1, j_2) \in \mathcal{J} \mid j_1 \leq j_1^* \text{ and } j_2 \leq j_2^*\}, \\
\mathcal{J}_2 &= \{(j_1, j_2) \in \mathcal{J} \mid j_1 m_1 \leq j_2 m_2 \text{ and } j_2 > j_2^*\}, \\
\mathcal{J}_3 &= \{(j_1, j_2) \in \mathcal{J} \mid j_1 m_1 > j_2 m_2 \text{ and } j_1 > j_1^*\}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\sum_{(j_1,j_2)\in\mathcal{J}_1} \sum_{k_1,k_2} \epsilon^2 \left(\frac{\lambda_{I,\epsilon}}{\epsilon} + 1\right) \varphi(\frac{\lambda_{I,\epsilon}}{\epsilon}) + \min\{\lambda_{I,\epsilon}^2, \theta_I^2\} \\
&= O\left(\epsilon^2 2^{j_1^*+j_2^*}\right) = O\left(\epsilon^{2\vartheta(m_1,m_2)}\right). \qquad (4.5)
\end{aligned}
$$

Further,

$$\begin{aligned}
&\sum_{(j_1,j_2)\in\mathcal{J}_2}\sum_{k_1,k_2}\epsilon^2\left(\frac{\lambda_{I,\epsilon}}{\epsilon}+1\right)\varphi(\frac{\lambda_{I,\epsilon}}{\epsilon})\\
&= \sum_{j_2>j_2^*}\sum_{j_1:j_1m_1\le j_2m_2} O\left(2^{j_1+j_2}\epsilon^2\sqrt{j_2-j_2^*}\exp\left\{-\frac{K^2_{m_1,m_2}(j_2-j_2^*)}{2m_1}\right\}\right)\\
&= \epsilon^2 2^{j_2^*(m_1+m_2)/m_1}\sum_{j_2>j_2^*} O\left(\exp\left\{\left(\log(2)(m_1+m_2)-K^2_{m_1,m_2}/2\right)(j_2-j_2^*)/m_1\right\}\sqrt{j_2-j_2^*}\right)\\
&= O\left(\epsilon^2 2^{j_2^*(m_1+m_2)/m_1}\right) = O\left(\epsilon^{2\vartheta(m_1,m_2)}\right). \qquad (4.6)
\end{aligned}$$

Here the last but one equality follows due to the convergence of the geometric series. Let $(j_1,j_2)\in\mathcal{J}_2$ be fixed. We choose $p=1$ if $p_2=1$ or $1<p<2$, $p\le p_2$ if $p_2>1$. By (4.4) we obtain

$$\#\left\{(k_1,k_2)\,\middle|\,|\theta_{(j_1j_2k_1k_2)}|>\lambda_{I,\epsilon}\right\} = O\left(\lambda_{I,\epsilon}^{-p}2^{-j_2m_2p}2^{(j_1+j_2)(1-p/2)}\right),$$

which implies that

$$\begin{aligned}
&\sum_{k_1,k_2}\min\left\{\lambda^2_{I,\epsilon},\theta_I^2\right\}\\
&= \epsilon^2\#\left\{(k_1,k_2)\,\middle|\,|\theta_{(j_1j_2k_1k_2)}|>\lambda_{I,\epsilon}\right\} + \sum_{(k_1,k_2):|\theta_{(j_1j_2k_1k_2)}|\le\lambda_{I,\epsilon}}\lambda^2_{I,\epsilon}\\
&= O\left(\lambda_{I,\epsilon}^{2-p}2^{-j_2m_2p}2^{(j_1+j_2)(1-p/2)}\right)\\
&= O\left(\epsilon^{2-p}(j_2-j_2^*)^{1-p/2}2^{-j_2m_2p}2^{(j_1+j_2)(1-p/2)}\right).
\end{aligned}$$

By $m_2>1/p$ we obtain that $[m_1m_2+(m_1+m_2)(p/2-1)]>0$, which yields

$$\begin{aligned}
&\sum_{(j_1,j_2)\in\mathcal{J}_2}\sum_{k_1,k_2}\min\left\{\lambda^2_{I,\epsilon},\theta_I^2\right\}\\
&= \epsilon^{2-p}2^{-j_2^*m_2p}\sum_{j_2>j_2^*}\sum_{j_1:j_1m_1\le j_2m_2} O\left((j_2-j_2^*)^{1-p/2}2^{(j_2^*-j_2)m_2p}2^{(j_1+j_2)(1-p/2)}\right)\\
&= \epsilon^{2-p}2^{-j_2^*(m_2-\frac{m_1+m_2}{2m_1})p}\sum_{j_2>j_2^*} O\left((j_2-j_2^*)^{1-p/2}2^{(j_2^*-j_2)[m_1m_2+(m_1+m_2)(p/2-1)]/m_1}\right)\\
&= O\left(\epsilon^{2\vartheta(m_1,m_2)}\right). \qquad (4.7)
\end{aligned}$$

The sum over $\mathcal{J}_3$ can be treated analogously to (4.6) and (4.7), which finishes the proof. $\square$

Proof of Lemma 2.3. It is easy to see that

$$\left\|f - Proj_{\widetilde{V}_J}f\right\|^2 = \sum_{j_1+j_2>J-2}\sum_{k_1,k_2}\theta_I^2 \le \sum_{(j_1,j_2)\in\mathcal{J}_4}\sum_{k_1,k_2}\theta_I^2 + \sum_{(j_1,j_2)\in\mathcal{J}_5}\sum_{k_1,k_2}\theta_I^2,$$

where

$$\begin{aligned} \mathcal{J}_4 &= \{(j_1, j_2) \mid L_1 j_1 \geq L_2 j_2 \quad \text{and} \quad j_1 > (J-2)L_2/(L_1+L_2)\}, \\ \mathcal{J}_5 &= \{(j_1, j_2) \mid L_1 j_1 < L_2 j_2 \quad \text{and} \quad j_2 > (J-2)L_1/(L_1+L_2)\} \end{aligned}$$

with

$$L_1 = m_1 - 1/\widetilde{p}_1 + 1/\widetilde{p}_2, \quad L_2 = m_2 - 1/\widetilde{p}_2 + 1/\widetilde{p}_1.$$

For the sake of a clear presentation we introduce the following notation

$$\theta_{(\psi,j_1,k_1),(\phi,j_2,k_2)} = \iint \psi_{j_1 k_1}(x_1)\phi_{j_2 k_2}(x_2) f(x_1, x_2)\, dx_1\, dx_2. \tag{4.8}$$

Now we get by Parseval's equality, Jensen's inequality and (4.3) that

$$\begin{aligned} &\sum_{j_2 :\, j_2 \leq j_1 L_1/L_2} \sum_{k_1,k_2} \theta_I^2 \\ &= \left\| Proj_{(W_{j_1} \otimes V_{[j_1 L_1/L_2+1]})} f \right\|^2 \\ &= \sum_{k_1,k_2} \theta^2_{(\psi,j_1,k_1),(\phi,[j_1 L_1/L_2+1],k_2)} \\ &\leq \left(\sum_{k_1,k_2} \left| \theta_{(\psi,j_1,k_1),(\phi,[j_1 L_1/L_2+1],k_2)} \right|^{\widetilde{p}_1} \right)^{2/\widetilde{p}_1} \\ &= O\left(2^{-2j_1 m_1} 2^{(j_1 + j_1 L_1/L_2)(2/\widetilde{p}_1 - 1)}\right) \\ &= O\left(2^{-j_1[2m_1 m_2 + m_1 + m_2 - 2m_1/\widetilde{p}_2 - 2m_2/\widetilde{p}_1]/L_2}\right). \end{aligned}$$

Since $[2m_1 m_2 + m_1 + m_2 - 2m_1/\widetilde{p}_2 - 2m_2/\widetilde{p}_1] > 0$, we have

$$\sum_{(j_1,j_2)\in\mathcal{J}_4} \sum_{k_1,k_2} \theta_I^2 = O\left(2^{-J[2m_1 m_2 + m_1 + m_2 - 2m_1/\widetilde{p}_2 - 2m_2/\widetilde{p}_1]/(L_1+L_2)}\right) = O\left(2^{-J\gamma(m_1,m_2,p_1,p_2)}\right).$$

The sum over $\mathcal{J}_5$ can be treated analogously, which proves the assertion. □

Proof of Theorem 2.3. Using (2.10), Lemma 2.2 and Lemma 2.3 we obtain, with $\delta^2 = \epsilon^2 \log(1/\epsilon)$, that

$$\begin{aligned} &\mathbb{E}\|\widehat{\widehat{f}} - f\|^2 \\ &= \sum_{I\in\mathcal{K}_\epsilon} \mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_\epsilon) - \theta_I\right)^2 + \sum_{I\notin\mathcal{K}_\epsilon} \theta_I^2 \\ &= O\left(\#\mathcal{K}_\epsilon \epsilon^2 \left(\sqrt{2\log(\#\mathcal{K}_\epsilon)} + 1\right) \varphi(\sqrt{2\log(\#\mathcal{K}_\epsilon)}) + \sum_{I\in\mathcal{K}_\epsilon} \min\{\delta^2, \theta_I^2\}\right) \\ &\quad + O\left(2^{-J_\epsilon \gamma(m_1,m_2,p_1,p_2)}\right) \\ &= O\left(\left(\epsilon^2\sqrt{2\log(\#\mathcal{K}_\epsilon)} + \Omega_\delta(\mathcal{B}, \mathcal{F}^{m_1,m_2}_{p_1,p_2})\right) + \epsilon^{2\vartheta(m_1,m_2)}\right) \\ &= O\left((\epsilon^2 \log(1/\epsilon))^{\vartheta(m_1,m_2)}\right). \end{aligned}$$

□

Proof of Lemma 2.4. Let, w.l.o.g., $m_1 \leq m_2$. Let j^* be such that $2^{j^*} \leq C_0 \epsilon^{-1/(m_1+1)} < 2^{j^*+1}$. To get a lower bound for $\Omega_\epsilon(\overline{\mathcal{B}}, \mathcal{F}_{p_1,p_2}^{m_1,m_2})$ consider the function

$$f_{\epsilon,1}(x_1,x_2) \;=\; \epsilon \sum_{k_1} 2^{j^*/2} \psi_{j^*,k_1}(x_1).$$

We have

$$\begin{aligned} \|f_{\epsilon,1}\|_{p_i} \;&\leq\; \|f_{\epsilon,1}\|_\infty \;\leq\; C\epsilon 2^{j^*} \;\leq\; CC_0 \epsilon^{m_1/(m_1+1)}, \\ \|\frac{\partial^{m_1}}{\partial x_1^{m_1}} f_{\epsilon,1}\|_{p_i} \;&\leq\; \|\frac{\partial^{m_1}}{\partial x_1^{m_1}} f_{\epsilon,1}\|_\infty \;\leq\; C\epsilon 2^{j^*(m_1+1)} \;\leq\; CC_0 \end{aligned}$$

and

$$\frac{\partial^{m_2}}{\partial x_2^{m_2}} f_{\epsilon,1} \equiv 0,$$

which implies that $f_{\epsilon,1} \in \mathcal{F}_{p_1,p_2}^{m_1,m_2}$ for an appropriate choice of C_0. With the exception of a negligible number of boundary wavelets we have, using notation (4.8),

$$\theta_{(\psi,j^*,k_1),(\phi,j^*,k_2)} \;=\; C\epsilon,$$

which implies that

$$\sum_{k_1,k_2} \min\left\{\epsilon^2, \theta^2_{(\psi,j^*,k_1),(\phi,j^*,k_2)}\right\} \;\geq\; C\epsilon^{2\overline{\vartheta}(m_1,m_2)}. \tag{4.9}$$

Let now $f \in \mathcal{F}_{p_1,p_2}^{m_1,m_2}$ be arbitrary. Then we have

$$\sum_{l-1\leq j\leq j^*} \sum_{k_1,k_2} \min\left\{\epsilon^2, \theta^2_{(\psi,j^*,k_1),(\phi,j^*,k_2)}\right\} \;=\; O\left(\epsilon^2 2^{2j^*}\right) \;=\; O\left(\epsilon^{2\overline{\vartheta}(m_1,m_2)}\right). \tag{4.10}$$

By (4.3) we get

$$\#\left\{(k_1,k_2) \,\Big|\, |\theta_{(\psi,j,k_1),(\phi,j,k_2)}| > \epsilon\right\} \;=\; O\left(\epsilon^{-\widetilde{p}_1} 2^{-j[(m_1+1)\widetilde{p}_1 - 2]}\right),$$

which, by $(m_1+1)\widetilde{p}_1 > 2$, implies that

$$\begin{aligned} &\sum_{j>j^*} \sum_{k_1,k_2} \min\left\{\epsilon^2, \theta^2_{(\psi,j,k_1),(\phi,j,k_2)}\right\} \\ &= \sum_{j>j^*} \epsilon^2 \#\left\{(k_1,k_2) \,\Big|\, |\theta_{(\psi,j,k_1),(\phi,j,k_2)}| > \epsilon\right\} + \sum_{j>j^*} \sum_{(k_1,k_2):|\theta_{(\psi,j,k_1),(\phi,j,k_2)}|\leq\epsilon} \theta^2_{(\psi,j,k_1),(\phi,j,k_2)} \\ &= \sum_{j>j^*} O\left(\epsilon^{2-\widetilde{p}_1} 2^{-j[(m_1+1)\widetilde{p}_1-2]}\right) \\ &= O\left(\epsilon^{2-\widetilde{p}_1} 2^{-j^*[(m_1+1)\widetilde{p}_1-2]}\right) \;=\; O\left(\epsilon^{2\overline{\vartheta}(m_1,m_2)}\right). \end{aligned} \tag{4.11}$$

The terms corresponding to the basis functions $\phi_{jk_1}(x_1)\psi_{jk_2}(x_2)$ as well as to $\psi_{jk_1}(x_1)\psi_{jk_2}(x_2)$ can be treated analogously. $\square$

Proof of Theorem 3.1. Let $g(u,\omega) := |A(u,\omega)|^2$. Neglecting the factor $1/(2\pi)$ we show that

$$R_T := \int_0^1 \int_{-\pi}^{\pi} \{ \sum_{s=-\infty}^{\infty} c_T(u,s)\ \exp(-i\omega s) \ - \ g(u,\omega)\}^2\, du\ d\omega \longrightarrow 0 \quad \text{as} \quad T \to \infty,$$

where

$$c_T(u,s) := cov\{X_{[uT-s/2],T}; X_{[uT+s/2],T}\} \ = \ \int_{-\pi}^{\pi} A([uT-s/2]/T,\lambda)\, \overline{A([uT+s/2]/T,\lambda)}\ \exp(i\lambda s)\, d\lambda.$$

Using the relation $\sum_s \exp(i(\lambda-\omega)s) = \delta(\lambda-\omega)$ we obtain

$$R_T = \int_0^1 \int_{-\pi}^{\pi} \{ \sum_{s=-\infty}^{\infty} \int [A([uT-s/2]/T,\lambda)\, \overline{A([uT+s/2]/T,\lambda)} - g(u,\lambda)]\ \exp(i(\lambda-\omega)s)\, d\lambda\}^2\, du\, d\omega.$$

Proceeding quite similarly as in the proof of Lemma 3.2(i) (on the rate of the bias), we have to estimate two terms of similar form. Hence, we only treat the first one which is

$$\widehat{\Delta}_s(u,s) = \int \Delta_s(u,\lambda)\ \exp(i\lambda s)\ d\lambda\,,$$

with

$$\Delta_s(u,\lambda) := \{A([uT-s/2]/T,\lambda) - A(u,\lambda)\}\, \overline{A(u,\lambda)}\ :$$

$$\begin{aligned}
&\int\int \left| \sum_s \widehat{\Delta}_s(u,s)\ \exp(-i\omega s)\right|^2\ d\omega\, du \\
&= \ \int du \sum_s \sum_v \widehat{\Delta}_s(u,s)\ \overline{\widehat{\Delta}_v(u,v)}\ \int \exp(-i\omega(s-v))\, d\omega \\
&= \ \sum_s \int_0^1 du\ |\widehat{\Delta}_s(u,s)|^2 \ = \ R_T^{(1)} \ + R_T^{(2)}\ ,
\end{aligned}$$

with

$$R_T^{(1)} \ = \ \sum_{|s|\le 2T} \ \sum_{n=1}^{[s_T^{-1}]+1} \int_{(n-1)s_T}^{(ns_T)\wedge 1} du\ |\widehat{\Delta}_s(u,s)|^2$$

and

$$R_T^{(2)} \ = \ \sum_{|s|>2T} \int_0^1 du\ |\widehat{\Delta}_s(u,s)|^2\ ,$$

where $s_T := |s|/(2T),\ 0 \le |s| \le 2T$.
Similarly to the proof of Lemma 3.2(i) we can show that

$$\begin{aligned}
&|s| \sup_{u\in[(n-1)s_T, ns_T]} |\widehat{\Delta}_s(u,s)| \\
&\le \ C_1 \left[\sup_{u,\lambda} |\overline{A(u,\lambda)}| + \sup_u\ TV_{[-\pi,\pi]}\,(\overline{A(u,\cdot)})\right]\ \sup_\lambda\ TV_{I_n(s_T)}\,(A(\cdot,\lambda)) \\
&\quad +\ C_2\ \sup_{u,\lambda}\{|\overline{A(u,\lambda)}|\,\}\, TV_{[-\pi,\pi]}\ \{TV_{I_n(s_T)}(A(\cdot,\cdot))\}\,,
\end{aligned}$$

where C_1 and C_2 denote some positive constants and where, on the right-hand side, the $\sup_u$ is taken over $u \in I_n(s_T) := [(n-1)s_T - 1/T, (n+1)s_T]$ and the $\sup_\lambda$ over $\lambda \in [-\pi, \pi]$.
Note that

$$\sum_{n=1}^{[s_T^{-1}]+1} \sup_{u \in I_n(s_T)} |\widehat{\Delta}_s(u,s)| \;=\; O(|s|^{-1}),$$

due to (A2) (a),(b),(c) (as $\sum_n TV_{I_n(s_T)}\,(A(\cdot,\lambda)) \le TV_{[0,1]}\,(A(\cdot,\lambda))$) .
Hence,

$$\begin{aligned}
|R_T^{(1)}| \;&\le\; \sum_{|s|\le 2T} |s_T| \sum_{n=1}^{[s_T^{-1}]+1} \sup_{u\in I_n(s_T)} |\widehat{\Delta}_s(u,s)|^2 \\
&\le\; (2T)^{-1} \sum_{|s|\le 2T} |s| \;\left\{ \sum_{n=1}^{[s_T^{-1}]+1} \sup_{u\in I_n(s_T)} |\widehat{\Delta}_s(u,s)| \right\}^2 \;=\; O(T^{-1}\log(T)).
\end{aligned}$$

Further,

$$|R_T^{(2)}| \;=\; O\left(\sum_{|s|>2T} s^{-2} \right) \;=\; O(T^{-1}),$$

as, by Definition (3.3) for $|s| > 2T$, $\Delta_s(u,\lambda) = \{A(0,\lambda)\,\overline{A(1,\lambda)} - |A(u,\lambda)|^2\}$ independent of s, hence, $\sup_{0\le u\le 1} |\widehat{\Delta}_s(u,s)| = O(|s|^{-1})$.
□

Proof of Lemma 3.2.
(i) We show that $R_T := |\mathbb{E}\,\widetilde{\theta}_I - \theta_I| = O(2^{(j_1+j_2)/2}T^{-1}\log(T))$. By (3.1) and with $A_t(\lambda) := A(t/T,\lambda)$, neglecting the factor $1/2\pi$,

$$\begin{aligned}
\mathbb{E}\, I_{t,T}(\omega) \;&=\; \sum_{|s/2|\le \min\{t-1,T-t\}} cov(X_{[t-s/2],T}; X_{[t+s/2],T}) \exp(-i\omega s) \\
&=\; \sum_{|s/2|\le \min\{t-1,T-t\}} \int_{-\pi}^{\pi} A_{[t-s/2]}(\lambda)\,\overline{A_{[t+s/2]}(\lambda)}\; \exp(i(\lambda-\omega)s)\; d\lambda\,.
\end{aligned}$$

Let $t_T := \min\{t-1, T-t\}$. According to the decomposition

$$\begin{aligned}
&A_{[t-s/2]}(\lambda)\overline{A_{[t+s/2]}(\lambda)} \;-\; A_t(\lambda)\overline{A_t(\lambda)} \\
&\quad =\; [A_{[t-s/2]}(\lambda) - A_t(\lambda)]\,\overline{A_t(\lambda)} \;+\; A_{[t-s/2]}(\lambda)\,[\overline{A_{[t+s/2]}(\lambda)} - \overline{A_t(\lambda)}]
\end{aligned}$$

we have $R_T = R_T^{(1)} + R_T^{(2)}$ with

$$\begin{aligned}
R_T^{(1)} \;&=\; \sum_t \int_{(t-1)/T}^{t/T} du\; \psi_{j_1k_1}(u) \sum_{s=-\infty}^{\infty} \int_{-\pi}^{\pi} d\omega\; \widetilde{\psi}_{j_2k_2}(\omega) \int_{-\pi}^{\pi} \{|A_t(\lambda)|^2 - f(u,\lambda)\} \exp(i(\lambda-\omega)s)\; d\lambda \\
&\quad - \sum_t \int_{(t-1)/T}^{t/T} du\; \psi_{j_1k_1}(u) \sum_{|s/2|>t_T} \int d\omega\, \widetilde{\psi}_{j_2k_2}(\omega)\; \exp(-i\omega s) \int |A_t(\lambda)|^2 \exp(i\lambda s)\; d\lambda \\
&=\; R_T^{(1,1)} + R_T^{(1,2)},
\end{aligned}$$

and

$$R_T^{(2)} = \sum_t \int_{(t-1)/T}^{t/T} du\ \psi_{j_1k_1}(u) \sum_{|s/2|\le t_T} \int d\omega\, \widetilde{\psi}_{j_2k_2}(\omega)\, \exp(-i\omega s) *$$
$$* \int \overline{A_t(\lambda)}\ [A_{[t-s/2]}(\lambda) - A_t(\lambda)]\ \exp(i\lambda s)\ d\lambda$$

where with $R_T^{(2)}$ we only treat the first part of two similar differences, w.l.o.g. Now, by (A2)(b), (A3) and (A4) b),

$$\begin{aligned} |R_T^{(1,1)}| &\le \int d\omega\, |\widetilde{\psi}_{j_2k_2}(\omega)| \sum_t \int_{(t-1)/T}^{t/T} du\ |\psi_{j_1k_1}(u)|\ TV_{[(t-1)/T,t/T]}(f(\cdot,\omega)) \\ &\le 2^{-j_2/2}2^{j_1/2}T^{-1} \sup_\omega TV_{[0,1]}(f(\cdot,\omega)) \\ &= O(2^{-j_2/2}2^{j_1/2}T^{-1})\ , \end{aligned}$$

and

$$\begin{aligned} |R_T^{(1,2)}| &\le \sup_u\{|\psi_{j_1k_1}(u)|\}T^{-1} \sum_t \sum_{|s/2|>t_T} |\widehat{\widetilde{\psi}_{j_2k_2}}(s)|\ \sup_t |\widehat{f}(t/T,s)| \\ &= O(2^{j_1/2}T^{-1}) \sum_t \sum_{|s/2|>t_T} O(2^{j_2/2}s^{-2}) \\ &= O(2^{(j_1+j_2)/2}T^{-1}\log(T))\ . \end{aligned}$$

Further, with s being even, w.l.o.g.,

$$R_T^{(2)} = -\sum_s \widehat{\widetilde{\psi}_{j_2k_2}}(s) \sum_t \int_{(t-1)/T}^{t/T} du\, \psi_{j_1k_1}(u) \int \overline{A_t(\lambda)} \sum_{n=0}^{s/2-1} \{A_{t-n}(\lambda) - A_{t-n-1}(\lambda)\} \exp(i\lambda s)\, d\lambda$$

such that, by (A2)(a), (b), (c), for some positive constant C,

$$\begin{aligned} |R_T^{(2)}| &\le C \sum_s |\widehat{\widetilde{\psi}_{j_2k_2}}(s)| \sup_u |\psi_{j_1k_1}(u)|\ T^{-1}\ |s|/2\ |s|^{-1} \Big[\sup_{u,\lambda}\{|A(u,\lambda)|\}\ \sup_\lambda\ TV_{[0,1]}\ (A(\cdot,\lambda)) + \\ &\qquad + \sup_{u,\lambda}\{|\overline{A(u,\lambda)}|\}\ TV_{U\times\Pi}(A) + \sup_u\ TV_{[-\pi,\pi]}\ (A(u,\cdot))\ \sup_\lambda\ TV_{[0,1]}\ (A(\cdot,\lambda)) \Big] \\ &= O(2^{j_1/2}2^{j_2/2}T^{-1}). \end{aligned}$$

The proof of the last estimate (for $R_T^{(2)}$) is delivered by some lengthy, but straightforward algebra using elementary generalizations of total variation estimates and partial summation. Roughly speaking, we proceed as follows: The integral w.r.t. λ delivers $s/2$ terms which are all of order $O(s^{-1})$, as for each of the differences labeled by n we use estimates like (cf. Edwards (1979), p. 34f.)

$$\int \Delta_t(\lambda)\ \exp(i\lambda s)\ d\lambda \sim \sum_k \Delta_t(\lambda_k)\{g_s(\lambda_k) - g_s(\lambda_{k-1})\} \sim -\sum_k \{\Delta_t(\lambda_{k+1}) - \Delta_t(\lambda_k)\} g_s(\lambda_k)$$

with $\Delta_t(\lambda) := \overline{A_t(\lambda)}\ (A_t(\lambda) - A_{t-1}(\lambda))$, $g_s(\lambda) := \exp(i\lambda s)/(is)$ and with a sufficiently fine partition $(\lambda_k)_k$ of $[-\pi,\pi]$. Note that $g_s(\lambda) = O(s^{-1})$.

The sum over t can be bounded from above by the bounded total variation of $\Delta_t(\lambda)$ as a function of u. Putting both (simultaneously) together, in order to strictly bound all occurring terms, we need Assumptions (A2)(a), (b) and (c), as $\Delta_t(\lambda)$ is a product of two functions of time and frequency.

(ii) To apply cumulant techniques we write $\widetilde{\theta}_I$ as a quadratic form with a symmetric matrix N_I:

$$\widetilde{\theta}_I \;=\; \underline{X}' N_I \underline{X},$$

where, $N_I = (M_I + \overline{M}_I)/2$ and, with $w_{j_1k_1}(t/T) \;:=\; T\; \int_{(t-1)/T}^{t/T} \psi_{j_1k_1}(u)\,du$ and $\widetilde{w}_{j_2k_2}(s) \;:=\; \widehat{\widetilde{\psi}_{j_2k_2}}(s) \;=\; (2\pi)^{-1}\; \int_{-\pi}^{\pi} \widetilde{\psi}_{j_2k_2}(\omega)\exp(-i\omega s)\,d\omega$,

$$(M_I)_{tv} \;=\; \begin{cases} T^{-1}\; w_{j_1k_1}(\frac{t+v}{2T})\widetilde{w}_{j_2k_2}(t-v) & \text{if } t+v \text{ even} \\ T^{-1}\; w_{j_1k_1}(\frac{t+v+1}{2T})\widetilde{w}_{j_2k_2}(t-v) & \text{if } t+v \text{ odd} \end{cases} .$$

In the following, for reasons of notational convenience, we use $w_{j_1k_1}(\frac{s+t}{2T})$ to denote $w_{j_1k_1}(\frac{[(s+t+1)/2]}{T})$. Note that, by the approximations used in the course of the proof, this does not lead to any problems.

Since $\widetilde{\phi}$ and $\widetilde{\psi}$ are of bounded variation, we get by integration by parts

$$\widetilde{w}_{j_2k_2}(t-v) \;=\; O\left(2^{-j_2/2} \wedge (2^{j_2/2}|t-v|^{-1})\right),$$

which implies that

$$(N_I)_{tv} \;=\; O\left(T^{-1}2^{j_1/2}\left[2^{-j_2/2} \wedge (2^{j_2/2}|t-v|^{-1})\right]\right).$$

Hence, we obtain the estimates

$$\max_{t,v}\{|(N_I)_{tv}|\} \;=\; O\left(T^{-1}2^{j_1/2}2^{-j_2/2}\right),$$

$$\|N_I\| \;\le\; \|N_I\|_\infty \;=\; O\left(T^{-1}2^{j_1/2}2^{j_2/2}\log(T)\right)$$

and

$$\widetilde{N}_I \;=\; \sum_s \max_t\{|(N_I)_{st}|\} \;=\; O\left(2^{-j_1/2}2^{-j_2/2}\right).$$

Let $\underline{Y} \sim N(0, Cov(\underline{X}))$. Since

$$\max_{t,v}\{|(N_I)_{tv}|\}\widetilde{N}_I \;=\; O\left(T^{-1}2^{-j_2}\right),$$

we obtain by Lemma 3.1 that

$$var(\widetilde{\theta}_I) \;=\; var(\underline{Y}'N_I\underline{Y}) \;+\; O(2^{-j_2}T^{-1}). \tag{4.12}$$

Now, with

$$\begin{aligned} var(\underline{Y}'N_I\underline{Y}) &= 2\;\; tr(N_I\Sigma_T N_I\Sigma_T) \qquad (4.13)\\ &= \;1/2\;[\;tr(M_I\Sigma_T M_I\Sigma_T) + tr(\overline{M}_I\Sigma_T\overline{M}_I\Sigma_T) + 2\;\; tr(\overline{M}_I\Sigma_T M_I\Sigma_T)\;], \end{aligned}$$

we have to show that

$$tr(\overline{M}_I \Sigma_T M_I \Sigma_T)$$
$$= \; 2\pi\; T^{-1} \int_{U\times\Pi} \{f(u,\omega)\,\psi_{j_1k_1}(u)\}^2\, du\; \tilde{\psi}_{j_2k_2}(\omega)\; \tilde{\psi}_{j_2k_2}(\omega)\, d\omega \;+\; o(T^{-1})$$

and

$$tr(M_I \Sigma_T M_I \Sigma_T) \;=\; tr(\overline{M}_I \Sigma_T \overline{M}_I \Sigma_T)$$
$$= \; 2\pi\; T^{-1} \int_{U\times\Pi} \{f(u,\omega)\,\psi_{j_1k_1}(u)\}^2\, du\; \tilde{\psi}_{j_2k_2}(\omega)\; \tilde{\psi}_{j_2k_2}(-\omega)\, d\omega \;+\; o(T^{-1})\;.$$

As this runs quite analogously for all terms under consideration, we restrict to treat the first one, only:

$$tr(\overline{M}_I \Sigma_T M_I \Sigma_T)$$
$$= \sum_{s,v,w,t} (\overline{M}_I)_{vw} (\Sigma_T)_{ws} (M_I)_{st} (\Sigma_T)_{tv}$$
$$= T^{-2} \sum_{s=1}^{T}\sum_{v=1}^{T}\sum_{w=1}^{T} w_{j_1k_1}(\frac{w+v}{2T})\, \overline{\tilde{w}_{j_2k_2}}(w-v)\; c_T(\frac{s+w}{2T}, s-w)$$
$$\sum_{t=1}^{T} w_{j_1k_1}(\frac{s+t}{2T})\; \tilde{w}_{j_2k_2}(s-t) c_T(\frac{t+v}{2T}, t-v)\,,$$

where we use the convention given in the beginning of our proof which allows to proceed regardless to the parity of the arguments of $w_{j_1k_1}$,
and where $\Sigma_T \;=\; Cov(\underline{X}) \;=\; (c_T(\cdot,\cdot))$ with

$$c_T(\frac{t}{T}, n) = cov\;\{X_{[t-n/2],T};\, X_{[t+n/2],T}\} = \int_{-\pi}^{\pi} A_{[t-n/2]}(\lambda)\; \overline{A_{[t+n/2]}(\lambda)}\exp(i\lambda n)\; d\lambda\,.$$

Note that with this, $c_T(\frac{t+v}{2T}, t-v) = cov\;\{X_{t,T};\, X_{v,T}\}$.
Further, let $\sum = (c(\cdot,\cdot))$ with $c(\frac{t}{T}, n) := (2\pi)^{-1}\int_{-\pi}^{\pi} f(\frac{t}{T},\lambda)\; \exp(i\lambda n)\, d\lambda$. For smooth A, $c_T(\frac{t}{T}, n) = c(\frac{t}{T}, n) \;+\; c'(\frac{\cdot}{T}, n)\; O(n/T)$ with both $\sup_{t/T}\sum_n |c(\frac{t}{T}, n)| < \infty$, and $\sup_{t/T}\sum_n |c'(\frac{t}{T}, n)| < \infty$.
If A is not smooth, but fulfills assumptions (A2) and (A3), then we proceed as in the proof of part (i) of this lemma, with the same quality of approximation (i.e. same resulting rates).
In the following, for sake of notational simplicity, we give the proof of (ii) only for functions $A(u,\lambda)$ and $\psi_{j_1k_1}(u)$, which are smooth in u.
To motivate the idea how to derive the leading term of the asymptotic variance, we briefly sketch the stationary situation (for details, cf. Gao (1993, page 19), but note the missing symmetrization of the Hermitian matrix M in that reference, which leads

to a slight mistake in the resulting asymptotic expression for the whole variance):

$$\begin{aligned} &T^{-2} \sum_{s,t,w,v} \overline{\tilde{w}_{jk}}(w-v)\ \tilde{w}_{jk}(s-t)\ c(s-w)\ c(t-v) \\ &= \ T^{-2} \sum_{l,m} \overline{\tilde{w}_{jk}}(l)\ \tilde{w}_{jk}(m) \sum_{s} c(s)\ c(s+l-m) + o(T^{-1}) \\ &= \ 2\pi T^{-1} \int_{-\pi}^{\pi} f^2(\omega)\ \tilde{\psi}_{jk}^2(\omega)\ d\omega + o(T^{-1}) \end{aligned}$$

(similarly to $\sum_{s=-T}^{T} \widehat{\tilde{\psi}_{jk}}(s)\ c(s) - \int \tilde{\psi}_{jk}(\omega)\ f(\omega)\ d\omega \ = \ \sum_{|s|>T} \widehat{\tilde{\psi}_{jk}}(s)\ c(s) \ = \ O(T^{-1})$).

To treat all of the occurring remainders, we use estimates like $\sum_n |\tilde{w}_{j_2k_2}(n)| \ = \ O(2^{j_2/2})$ (due to (A4) b)) and $\sum_n |c(u,n)| \ < \ \infty$ uniformly in $u \in [0,1]$ (due to (A3)).
Note in particular that $\sum_n |n|\ |l_1(n)|\ \ |l_2(n)| < \infty$, where $l_i(n)$ is any of $\tilde{w}(n), c(\cdot,n)$ or even of $\sum_w \tilde{w}(w-v)\ c(s-w) = l(s-v)$, say, with again $\sum_n |l(n)| < \infty$.

Our proof proceeds by three different approximations: The first is replacing $c_T(\frac{t}{T},n)$ by $c(\frac{t}{T},n)$ with an error of order $O(n/T)$ (see above). The second one is to replace $c(\frac{s+w}{2T},\cdot)$ by $c(\frac{s}{2T},\cdot) + O(w/T)$ and $w_{j_1k_1}(\frac{w+v}{2T})$ by $w_{j_1k_1}(\frac{v}{2T}) + O(2^{j_1/2}\ w/T)$. With this,

$$\begin{aligned} &tr(\overline{M}_I \Sigma M_I \Sigma) \\ &= \ T^{-2} \sum_s \sum_v \sum_w \left[\ w_{j_1k_1}(\frac{v}{2T}) \ + \ O(2^{j_1/2}\ \frac{w}{T})\ \right]\ \overline{\tilde{w}_{j_2k_2}}(w-v)\ \left[\ c(\frac{s}{2T}, s-w) \ + \ O(\frac{w}{T})\ \right] \\ &\quad \sum_t \left[\ w_{j_1k_1}(\frac{s}{2T}) \ + \ O(2^{j_1/2}\ \frac{t}{T})\ \right]\ \tilde{w}_{j_2k_2}(s-t)\ \left[\ c(\frac{v}{2T}, t-v) \ + \ O(\frac{t}{T})\ \right]. \end{aligned}$$

The leading term of $tr(\overline{M}_I \Sigma M_I \Sigma)$ turns out to be

$$T^{-2} \sum_s \sum_v w_{j_1k_1}(\frac{v}{2T})\, w_{j_1k_1}(\frac{s}{2T}) \int\!\!\int d\lambda\, d\tilde{\lambda}\, f(\frac{v}{2T},\lambda)\, f(\frac{s}{2T},\tilde{\lambda})\, \tilde{\psi}_{j_2k_2}(\lambda)\, \tilde{\psi}_{j_2k_2}(\tilde{\lambda})\ \exp(i(s-v)(\lambda-\tilde{\lambda})).$$

The occurring remainders of both first (i.e. replacing $c_T(\cdot,\cdot)$ by $c(\cdot,\cdot)$) and second approximation are of the following kind (or even of higher order):
$T^{-2} \sum_{s,v,w,t} w_{j_1k_1}(\frac{v}{2T})\ w_{j_1k_1}(\frac{s}{2T})\ \overline{\tilde{w}_{j_2k_2}}(w-v)\ \tilde{w}_{j_2k_2}(s-t)\ c(\frac{s}{2T}, s-w)\ c'(\frac{\cdot}{2T}, t-v)\ O(\frac{t}{T})$.

In each of these remainders use estimates like
$\sum_s \sum_t \frac{|t|}{T}\ |\tilde{w}_{j_2k_2}(s-t)|\ \ |c'(\frac{\cdot}{2T}, t-s)| = o(2^{j_2/2})$, and respectively,

$$\begin{aligned} &T^{-2} \sum_s \sum_v \sum_t |\ w_{j_1k_1}(\frac{v}{2T})\ w_{j_1k_1}(\frac{s}{2T})\ l(s-v)\ \ \frac{|t|}{T}\ \tilde{w}_{j_1k_1}(s-t)\ c'(\frac{\cdot}{2T}, t-s)\ | \\ &= \ O(2^{(j_1+j_2)}T^{-2}) \ = \ o(T^{-1}). \end{aligned}$$

Finally, the third approximation, which is

$$f(\frac{v}{2T},\tilde{\lambda}) = f(\frac{s}{2T},\tilde{\lambda}) + f'(\frac{\cdot}{T},\tilde{\lambda})\ O(\frac{s-v}{T}) \text{ and } w_{j_1k_1}(\frac{v}{2T}) = w_{j_1k_1}(\frac{s}{2T}) + w'_{j_1k_1}(\frac{\cdot}{T})\ O(2^{j/2}\frac{s-v}{T}),$$

delivers a leading term, with $n := s - v$,

$$T^{-2} \sum_s \sum_{|n| \le T} w_{j_1 k_1}^2 (\frac{s}{2T}) [\int \int d\lambda \, d\tilde{\lambda} \, \tilde{\psi}_{j_2 k_2}(\lambda) \, \tilde{\psi}_{j_2 k_2}(\tilde{\lambda}) \, f(\frac{s}{2T}, \lambda) \, f(\frac{s}{2T}, \tilde{\lambda}) \, \exp(in(\lambda - \tilde{\lambda})) + \\ + R_T^{(3)}(n) \,]$$

with

$$\sum_n |R_T^{(3)}(n)| \; = \; \sum_n \frac{|n|}{T} \, |\hat{\mathcal{F}}(\cdot, n)|^2 \; = \; O(2^{j_2} T^{-1}),$$

where $\hat{\mathcal{F}}(\cdot, n) = \int \tilde{\psi}_{j_2 k_2}(\lambda) \, f(\cdot, \lambda) \, \exp(in\lambda) \, d\lambda$ is again absolutely summable as a function of n, uniformly in its first argument, and with $T^{-2} \sum_s w_{j_1 k_1}^2 (\frac{s}{2T}) = O(2^{j_1} T^{-1})$.

We finish the proof by a technique similar to the proof of part (i), i.e., replacing $\sum_{|n| \le T} \ldots$ by $\sum_{n=-\infty}^{\infty} \ldots$, noting that $\sum_{|n| \ge cT} |\hat{\mathcal{F}}(\cdot, n)|^2 \; = \; O(T^{-1})$.
Hence we end up with the following overall leading term of $tr(\overline{M}_I \Sigma M_I \Sigma)$,

$$\begin{aligned} & 2\pi \; T^{-2} \sum_s w_{j_1 k_1}^2 (\frac{s}{2T}) \int d\lambda \, \tilde{\psi}_{j_2 k_2}^2(\lambda) \; f^2(\frac{s}{2T}, \lambda) \\ = \; & 2\pi \; T^{-1} \int_0^1 du \; \psi_{j_1 k_1}^2(u) \int_{-\pi}^{\pi} d\lambda \, \tilde{\psi}_{j_2 k_2}^2(\lambda) \, f^2(u, \lambda) \; + \; O(T^{-2} 2^{(j_1 + j_2)}), \end{aligned}$$

due to the bounded total variation of all occurring functions.
The proof of (ii) ends by applying the same techniques to the remaining two terms of the sum in (4.13).

(iii) This can be shown simply by using Lemma 3.1 with, by (A5),

$$\begin{aligned} \lambda_{\max} \left(M_I \, Cov(\underline{X}) \right) \; & \le \; \|M_I\| \, \| \, Cov(\underline{X}) \| \\ & = \; O\left(T^{-1} 2^{(j_1 + j_2)/2} \log(T) \right) \sup_{1 \le t \le T} \left\{ \sum_s cov(X_s, X_t) \right\} \\ & = \; O\left(T^{-1} 2^{(j_1 + j_2)/2} \log(T) \right) \end{aligned}$$

and the estimates for $\max_{u,v}\{|(M_I)_{uv}|\}$ and $\|M_I\|_\infty$ derived in the proof of (ii). $\square$

Proof of Proposition 3.1. By (ii) of Lemma 3.2 we get, in conjunction with (A2)c) and d), that $\sigma_I \asymp T^{-1/2}$ for $T^\rho \le 2^{j_2}$. Hence, we obtain by (iii) of Lemma 3.2, for appropriate $\mu > 0$,

$$| \, cum_n(\tilde{\theta}_I / \sigma_I)| \; \le \; (n!)^{2+2\gamma} \left(C T^{-1/2} 2^{(j_1 + j_2)/2} \log(T) \right)^{n-2} \; \le \; (n!)^{2+2\gamma} \, (C T^\mu)^{-(n-2)} \tag{4.14}$$

for all $n \ge 3$, which implies by Lemma 1 in Rudzkis, Saulis and Statulevicius (1978) that

$$P\left(\pm(\tilde{\theta}_I - \mathbb{E}\tilde{\theta}_I)/\sigma_I \ge x \right) \; = \; (1 \; - \; \Phi(x))(1 + o(1)) \tag{4.15}$$

holds uniformly in $0 \le x \le T^\vartheta$ for some $\vartheta > 0$.

With $\Delta_I := (\mathbb{E}\widetilde{\theta}_I - \theta_I)/\sigma_I = o(1)$ we get

$$P\left(\pm(\widetilde{\theta}_I - \theta_I)/\sigma_I \geq x\right) = (1 - \Phi(x))(1 + o(1)) + O\left(|\Phi(x) - \Phi(x + \Delta_I)|\right).$$

Fix any $c > 1$. For $x \leq c$, obviously

$$\Phi(x) - \Phi(x + \Delta_I) = o(1 - \Phi(x)). \tag{4.16}$$

Let w.l.o.g. $\Delta_I \geq 0$. Using the formula $(1 - 1/x^2)\varphi(x)/x \leq (1 - \Phi(x))$ we obtain for $x > c$ that

$$|\Phi(x) - \Phi(x + \Delta_I)| = \Delta_I \varphi(x) = o(1 - \Phi(x)), \tag{4.17}$$

which, in conjunction with (4.15) and (4.16), completes the proof. $\square$

Proof of Proposition 3.2. First, let $I \in \mathcal{I}_T \cap \{I \mid 2^{j_2} > T^\rho\}$. Since $\delta^{(\cdot)}$ is monotonic in its first argument, there exists some γ_I such that

$$\begin{aligned}
\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,T}) &\geq \theta_I, \quad \text{if} \quad \widetilde{\theta}_I - \theta_I > \gamma_I, \\
\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,T}) &\leq \theta_I, \quad \text{if} \quad \widetilde{\theta}_I - \theta_I < \gamma_I.
\end{aligned}$$

W.l.o.g. we assume that $\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,T}) \geq \theta_I$, if $\widetilde{\theta}_I - \theta_I = \gamma_I$.
Let $\eta_T = CT^{-1/2}\sqrt{\log(T)}$ for some appropriate C. Then

$$\begin{aligned}
&\mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,T}) - \theta_I\right)^2 \\
&= \mathbb{E}I\left(\gamma_I \leq \widetilde{\theta}_I - \theta_I < \eta_T\right)\left(\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,T}) - \theta_I\right)^2 \\
&\quad + \mathbb{E}I\left(-\eta_T < \widetilde{\theta}_I - \theta_I < \gamma_I\right)\left(\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,T}) - \theta_I\right)^2 \\
&\quad + \mathbb{E}I\left(|\widetilde{\theta}_I - \theta_I| \geq \eta_T\right)\left(\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,T}) - \theta_I\right)^2 \\
&= S_1 + S_2 + S_3.
\end{aligned}$$

Applying integration by parts w.r.t. x, we obtain by Proposition 3.1 that

$$\begin{aligned}
S_1 &= -\int \left[I\left(\gamma_I \leq \sigma_I x < \eta_T\right)\left(\delta^{(\cdot)}(\theta_I + \sigma_I x, \lambda_{I,T}) - \theta_I\right)^2\right] d\left\{P((\widetilde{\theta}_I - \theta_I)/\sigma_I \geq x)\right\} \\
&= \int \left\{P((\widetilde{\theta}_I - \theta_I)/\sigma_I \geq x)\right\} d\left[I\left(\gamma_I \leq \sigma_I x < \eta_T\right)\left(\delta^{(\cdot)}(\theta_I + \sigma_I x, \lambda_{I,T}) - \theta_I\right)^2\right] \\
&\quad + P((\widetilde{\theta}_I - \theta_I)/\sigma_I \geq \gamma_I)\left(\delta^{(\cdot)}(\theta_I + \gamma_I, \lambda_{I,T}) - \theta_I\right)^2 \\
&\leq C_T\left\{\int \{1 - \Phi(x)\}\, d\left[I\left(\gamma_I \leq \sigma_I x < \eta_T\right)\left(\delta^{(\cdot)}(\theta_I + \sigma_I x, \lambda_{I,T}) - \theta_I\right)^2\right]\right. \\
&\qquad\qquad \left. + P((\xi_I - \theta_I)/\sigma_I \geq \gamma_I)\left(\delta^{(\cdot)}(\theta_I + \gamma_I, \lambda_{I,T}) - \theta_I\right)^2\right\} \\
&= C_T \mathbb{E}I\left(\gamma_I \leq \xi_I - \theta_I < \eta_T\right)\left(\delta^{(\cdot)}(\xi_I, \lambda_{I,T}) - \theta_I\right)^2
\end{aligned} \tag{4.18}$$

holds uniformly in $I \in \mathcal{I}_T \cap \{I \mid 2^{j_2} > T^\rho\}$ for some $C_T \to 1$. The term S_2 can be estimated analogously.

Using Lemma 3.2 we obtain, for arbitrary even n, that

$$\mathbb{E}\left(\widetilde{\theta}_I - \theta_I\right)^n = O\left(\sum_{r=1}^{n} \prod_{i_1,\dots,i_r : i_1+\cdots+i_r=n, i_j\ge 1} |\, cum_{i_j}(\widetilde{\theta}_I)|\right) = O(T^{-n/2}),$$

which implies, by Cauchy-Schwarz, that

$$S_3 \le \sqrt{P\left(|\widetilde{\theta}_I - \theta_I| \ge \eta_T\right)}\sqrt{\mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,T}) - \theta_I\right)^4} = O(T^{-2}), \qquad (4.19)$$

if C is chosen large enough.
As $|\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,T}) - \theta_I| \le \lambda_{I,T} + |\widetilde{\theta}_I - \theta_I|$, the terms with $2^{j_2} \le T^{\rho}$ contribute to the risk a term of order $O(T^{\rho-1}\log(T))$, which is $O(T^{-\vartheta(m_1,m_2)})$, if ρ is chosen sufficiently small. □

Proof of Theorem 3.2. Using Parseval's identity we infer from Proposition 3.2 and by $|\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda) - \theta_I| \le \lambda + |\widetilde{\theta}_I - \theta_I|$ that

$$\begin{aligned}
\mathbb{E}\|\widehat{f} - f\|^2 &= \sum_{I\in\mathcal{I}_T} \mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,T}) - \theta_I\right)^2 + \sum_{I\notin\mathcal{I}_T} \theta_I^2 \\
&= (1 + o(1)) \sum_{I\in\mathcal{I}_T^*} \mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\xi}_I, \lambda_{I,T}) - \theta_I\right)^2 + O\left(T^{-\vartheta(m_1,m_2)}\right) \\
&\quad + \sum_{I\in\mathcal{I}_T\setminus\mathcal{I}_T^*} \left(2\lambda_{I,T}^2 + 2\mathbb{E}(\widetilde{\theta}_I - \theta_I)^2\right) \qquad (4.20) \\
&\quad + O\left(T^{-(1-\delta)\gamma(m_1,m_2,p_1,p_2)}\right).
\end{aligned}$$

From (2.10) we see that the first term on the right-hand side of (4.20) can be estimated by

$$\begin{aligned}
& C \sum_{I\in\mathcal{I}_T^*} \left(\sigma_I^2\left(\frac{\lambda_{I,T}}{\sigma_I} + 1\right)\varphi(\frac{\lambda_{I,T}}{\sigma_I}) + \min\{\lambda_{I,T}^2, \theta_I^2\}\right) + O\left(T^{-\vartheta(m_1,m_2)}\right) \\
&\le C \sum_{I\in\mathcal{I}_T^*} \min\{\sigma_I^2\log(T), \theta_I^2\} + O\left(T^{-\vartheta(m_1,m_2)}\right) \\
&= O\left(\Omega_{\max\{\sigma_I\sqrt{\log(T)}\}}(\mathcal{B}, \mathcal{F}_{p_1,p_2}^{m_1,m_2}) + T^{-\vartheta(m_1,m_2)}\right) \\
&= O\left((\log(T)/T)^{-\vartheta(m_1,m_2)}\right).
\end{aligned}$$

The remaining terms on the right-hand side of (4.20) are also of order $O(T^{-\vartheta(m_1,m_2)}\log(T))$, which finishes the proof. □

Proof of Theorem 3.3. Since $\delta^{(\cdot)}$ is monotonic in the second argument, we have for any random threshold $\widehat{\lambda}_I$ satisfying $\lambda_{I,1} \le \widehat{\lambda}_I \le \lambda_{I,2}$ that $|\delta^{(\cdot)}(\widetilde{\theta}_I, \widehat{\lambda}_I) - \theta_I| \le \max\{|\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,1}) - \theta_I|, |\delta^{(\cdot)}(\widetilde{\theta}_I, \lambda_{I,2}) - \theta_I|\}$. For $CT^{-1/2}\sqrt{\log(T)} \ge \gamma_T\sigma_I\sqrt{2\log(\#\mathcal{I}_T^0)}$, both the nonrandom thresholds $\lambda_{I,T} = \gamma_T \sigma_I \sqrt{2\log(\#\mathcal{I}_T^0)}$ as well as $\lambda_{I,T} = CT^{-1/2}\sqrt{\log(T)}$ provide the desired rate for the risk. Hence we obtain

$$\begin{aligned}
&\sum_{I\in\mathcal{I}_T^0} \mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\theta}_I,\widehat{\lambda}) - \theta_I\right)^2 \\
&\leq \sum_{I\in\mathcal{I}_T^0} \mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\theta}_I,\gamma_T\sigma_I\sqrt{2\log(\#\mathcal{I}_T^0)}) - \theta_I\right)^2 \\
&\quad + \sum_{I\in\mathcal{I}_T^0} \mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\theta}_I, CT^{-1/2}\sqrt{\log(T)}) - \theta_I\right)^2 \\
&\quad + \sum_{I\in\mathcal{I}_T^0} \mathbb{E} I\left(\widehat{\lambda} \notin [\gamma_T\sigma_I\sqrt{2\log(\#\mathcal{I}_T^0)}, CT^{-1/2}\sqrt{\log(T)}]\right)\left(2\widetilde{\theta}_I^2 + 2\theta_I^2\right) \\
&= O\left((\log(T)/T)^{-\vartheta(m_1,m_2)}\right).
\end{aligned}$$

From the proof of Theorem 3.2 we know that the risk arising from the estimation of θ_I, $I \notin \mathcal{I}_T^0$, is also of order $O\left((\log(T)/T)^{-\vartheta(m_1,m_2)}\right)$, which finishes the proof. □

Acknowledgement. We thank D. Donoho and P. Mathé for helpful discussions.

References

1. Berkolajko, M. Z. and Novikov, I. Ya. (1992). Bases of wavelets in spaces of differentiable functions of anisotropic smoothness. *Russian Acad. Sci. Dokl. Math.* **45**, 382–386.
2. Bretagnolle, J. and Huber, C. (1979). Estimation des densités: risque mimimax. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **47**, 119–137.
3. Brown, D. L. and Low, M. (1992). Asymptotic equivalence of nonparametric regression and white noise. Manuscript.
4. Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transform. *Appl. Comp. Harmonic Anal.* **1**, 54–81.
5. Dahlhaus, R. (1993). Fitting time series models to nonstationary processes. Preprint, University of Heidelberg.
6. Dahlhaus, R. (1994). On the Kullback-Leibler information divergence of locally stationary processes. Preprint, University of Heidelberg.
7. Daubechies, I. (1992). *Ten Lectures on Wavelets*, SIAM, Philadelphia.
8. Delyon, B. and Juditsky, A. (1993). Wavelet estimators, global error measures: revisited. Technical Report No. 782, Irisa, France.
9. Donoho, D. L. and Johnstone, I. M. (1994a). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455.
10. Donoho, D. L. and Johnstone, I. M. (1994b). Ideal denoising in an orthonormal basis chosen from a library of bases. Technical Report, Department of Statistics, Stanford University.
11. Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? *J. R. Statist. Soc., Ser. B* **57**, to appear.
12. Edwards, R. E. (1979). *Fourier Series. A Modern Introduction.*, 2. Ed., Vol. 1, Springer, New York.
13. Gao, H.-Y. (1993). Wavelet estimation of spectral densities in time series analysis. Ph. D. dissertation. U.C. Berkeley.

14. Meyer, Y. (1991). Ondelettes sur l'intervalle. *Revista Mathematica Ibero-Americana* **7** (2), 115–133.
15. Neumann, M. H. (1994). Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series, Manuscript.
16. Neumann, M. H. and Spokoiny, V. G. (1995). On the efficiency of wavelet estimators under arbitrary error distributions. *Math. Methods of Statist.* **4**, 137–166.
17. Nikol'skii, S. M. (1975). *Approximation of Functions of Several Variables and Imbedding Theorems.* Springer.
18. Nussbaum, M. (1994). Asymptotic equivalence of density estimation and white noise. to appear *Ann. Statist.*
19. Priestley, M. B. (1981). *Spectral Analysis and Time Series.* Vol. 2, Academic Press, London.
20. Rudzkis, R. (1978). Large deviations for estimates of spectrum of stationary series. *Lithuanian Math. J.* **18**, 214-226.
21. Rudzkis, R., Saulis, L. and Statulevicius, V. (1978). A general lemma on probabilities of large deviations. *Lithuanian Math. J.* **18**, 226-238.
22. von Sachs, R. and Schneider, K. (1994). Wavelet smoothing of evolutionary spectra by non-linear thresholding. Technical Report, University of Kaiserslautern.
23. Tribouley, K. (1995). Practical estimation of multivariate densities using wavelet methods. *Statistica Neerlandica* **49**, 41–62.
24. Walsh, J. B. (1986). Martingales with a multidimensional parameter and stochastic integrals in the plane. In *Lectures in Probability and Statistics* (A. Dold and B. Eckmann, ed.) *Lecture Notes in Math.* **1215**, 329-491. Springer, Berlin.